REFERENCES

Adeli, Adeli, E, Wu, G, Saghafi, B, An, L, Shi, F & Shen, D 2017, 'Kernel-based joint feature selection and max-margin classification for early diagnosis of Parkinson's disease,' Scientific reports, vol. 7, no. 1, pp. 1-14.

Aich, S, Younga, K, Hui, KL, Al-Absi, AA & Sain, M 2018, 'A nonlinear decision tree-based classification approach to predict the Parkinson's disease using different feature sets of voice data,' In 2018 20th International Conference on Advanced Communication Technology (ICACT) (pp. 638-642). IEEE..

Alemami Y & Almazaydeh, L 2014, 'Detection of Parkinson disease through voice signal features,' Journal of American Science, vol. 10, no. 10, pp. 44–47.

Al-Fatlawi, A.H.; Jabardi, M.H.; Ling, S.H. Efficient Diagnosis System for Parkinson's Disease Using Deep Belief Network. In Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 24–29 July 2016; pp. 1324–1330.

Alhussein, M 2017, 'Monitoring Parkinson's disease in smart cities,' IEEE Access, vol. 5, pp.19835-19841.

Ali, L, Zhu, C, Golilarz, NA, Javeed, A, Zhou, M & Liu, Y 2019, 'Reliable Parkinson's disease detection by analyzing handwritten drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model,' IEEE Access, vol. 7, pp. 116480-116489.

Ali, L, Zhu, C, Zhang, Z & Liu, Y 2019, 'Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network,' IEEE Journal of Translational Engineering in Health and Medicine, vol. 7, pp.1-10.

Ali, L, Zhu, C, Zhang, Z & Liu, Y 2019, ‘Automated detection of Parkinson’s disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network, ’IEEE Journal of Translational Engineering in Health and Medicine, vol. 7, pp. 1-10.

Ali, M, Ali, SI, Kim, D, Hur, T, Bang, J, Lee, S, Kang, BH &Hussain, M 2018, ‘uEFS: An efficient and comprehensive ensemble-based feature selection methodology to select informative features,’ PloS one, vol. 13, no. 8, pp. 1-28.

Alweshah, M 2020, ‘Solving feature selection problems by combining mutation and crossover operations with the monarch butterfly optimization algorithm,’ Applied Intelligence, pp. 1-24.

Alweshah, M, Khalaileh, SA, Gupta, BB, Almomani, A, Hammouri, AI & Al-Betar, MA, 2020, ‘The monarch butterfly optimization algorithm for solving feature selection problems,’ Neural Computing and Applications, pp. 1-15.

Alweshah, M. 2020, ‘Solving feature selection problems by combining mutation and crossover operations with the monarch butterfly optimization algorithm,’ Applied Intelligence, pp. 1-24.

Anisha, CD & Arulanand, N 2020, ‘Early Prediction of Parkinson's Disease (PD) Using Ensemble Classifiers,’ In 2020 International Conference on Innovative Trends in Information Technology (ICITIIT) (pp. 1-6). IEEE.

Arora, J, Agrawal, U, Tiwari, P, Gupta, D & Khanna, A 2020, ‘Ensemble feature selection method based on recently developed nature-inspired algorithms,’ In International Conference on Innovative Computing and Communications (pp. 457-470). Springer, Singapore.

Arora, S & Singh, S 2019, ‘Butterfly optimization algorithm: a novel approach for global optimization,’ Soft Computing, vol. 23, no. 3, pp. 715-734.

Ashour, AS, El-Attar, A, Dey, N, Abd El-Kader, H & Abd El-Naby, MM 2020, ‘Long Short-Term Memory based patient-dependent model for FOG detection in Parkinson's disease,’ Pattern recognition letters, vol. 131, pp. 23-29.

Awada, W, Khoshgoftaar, TM, Dittman, D, Wald, R & Napolitano, A 2012, ‘A review of the stability of feature selection techniques for bioinformatics data,’ In: IEEE 13th international conference on information reuse and integration, pp 356–363

Balaji, E, Brindha, D & Balakrishnan, R 2020, ‘Supervised machine learning based gait classification system for early detection and stage classification of Parkinson’s disease,’ Applied Soft Computing, vol. 94, pp. 106494.

Ball, N, Teo, WP, Chandra, S & Chapman, J 2019, ‘Parkinson's disease and the environment,’ Frontiers in neurology, vol. 10, pp. 218.

Basir, MA, Hussin, MS & Yusof, Y 2021, ‘Ensemble Feature Selection Method Based on Bio-inspired Algorithms for Multi-objective Classification Problem’, in Advances on Smart and Soft Computing (pp. 167-176). Springer, Singapore..

Benba, A, Jilbab, A, Hammouch, A & Sandabad, S 2015, ‘Voiceprints analysis using MFCC and SVM for detecting patients with Parkinson's disease,’ In 2015 International conference on electrical and information technologies (ICEIT), pp. 300-304.

Bhati, S, Velazquez, LM, Villalba, J & Dehak, N 2019, ‘LSTM- siamese network for Parkinson’s disease detection from speech,’ In 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp. 1-5). IEEE.

Bodini, M 2022, ‘Opinion mining from machine translated Bangla reviews with stacked contractive auto-encoders,’ Journal of Ambient Intelligence and Humanized Computing, pp. 1-13.

Boutsidis, C, Zouzias, A, Mahoney, MW & Drineas, P 2014, ‘Randomized dimensionality reduction for $ k $-means clustering,’ IEEE Transactions on Information Theory, vol. 61, no. 2, pp. 1045-1062.

Bühlmann, P 2012, ‘Bagging, boosting and ensemble methods,’ In Handbook of computational statistics (pp. 985-1022). Springer, Berlin, Heidelberg.

C. Quan, K. Ren, and Z. Luo, “A deep learning based method for Parkinson’s disease detection using dynamic features of speech,” IEEE Access, vol. 9, pp. 10239–10252, 2021.

Cai, Z, Gu, J & Chen, HL 2017, ‘A new hybrid intelligent framework for predicting Parkinson’s disease,’ IEEE Access, vol. 5, pp. 17188-17200.

Caliskan, A, Badem, H, Basturk, A &Yuksel, ME 2017, ‘Diagnosis of the Parkinson disease by using deep neural network classifier,’ IU-Journal of Electrical & Electronics Engineering, vol. 17, no. 2, pp. 3311-3318.

Cao, Y, Geddes, TA, Yang, JYH and Yang, P 2020, ‘Ensemble deep learning in bioinformatics,’ Nature Machine Intelligence, vol. 2, no. 9, pp. 500-508.

Chen T, Xu, R, He Y & Wang, X 2017, ‘Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,’ Expert Syst. Appl., vol. 72, pp. 221–230, Apr.

Chen, CW, Tsai, YH, Chang, FR & Lin, WC 2020, ‘Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results,’ Expert Systems, vol. 37, no. 5, pp. e12553.

Chen, S, Luo, Z & Gan, H 2018, ‘An entropy fusion method for feature extraction of EEG,’ Neural Computing and Applications, vol. 29, no. 10, pp. 857-863.

Chen, T, Xu, R, He, Y, Wang, X 2017, ‘Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,’ Expert Syst. Appl., vol. 72, pp. 221–230.

Chia, SJ, Tan, EK & Chao, YX 2020, ‘Historical perspective: Models of Parkinson’s disease,’ International journal of molecular sciences, vol. 21, no. 7, pp. 2464.

Cigdem, O & Demirel, H 2018, ‘Performance analysis of different classification algorithms using different feature selection methods on Parkinson's disease detection,’ Journal of Neuroscience Methods, 309, pp.81-90.

Cigdem, O, Yilmaz, A, Beheshti, I & Demirel, H 2018, ‘Comparing the performances of PDF and PCA on Parkinson's disease classification using structural MRI images,’ In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

Dong, X, Yu, Z, Cao, W, Shi, Y & Ma, Q 2020, ‘A survey on ensemble learning,’ Frontiers of Computer Science, vol. 14, no. 2, pp. 241-258.

Dubey, AK 2021, ‘Optimized hybrid learning for multi disease prediction enabled by lion with butterfly optimization algorithm,’ Sādhanā, vol. 46, no. 2, pp. 1-27.

El Maachi, I, Bilodeau, GA &Bouachir, W 2020, ‘Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait,’ Expert Systems with Applications, vol. 143, pp. 113075.

Eskofier, BM, Lee, SI, Daneault, JF, Golabchi, FN, Ferreira- Carvalho, G, Vergara-Diaz, G, Sapienza, S, Costante, G, Klucken, J, Kautz T 2016, ‘Recent machine learning advancements in sensor-based mobility analysis: Deep learning for Parkinson’s disease assessment,’ In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 655–658.

Fanning, S, Selkoe, D &Dettmer, U 2020, ‘Parkinson’s disease: proteinopathy or lipidopathy,’ NPJ Parkinson's disease, vol. 61, pp.1-9.

Faraji, B, Esfahani, Z, Rouhollahi, K & Khezri, D 2019, ‘Optimal Canceling of the Physiological Tremor for Rehabilitation in Parkinson’s disease,’ Journal of Exercise Science and Medicine, vol. 11, no. 2, pp. 113-124

Feng, Y, Wang G-G, Li W, Li N 2018, ‘Multi-strategy monarch butterfly optimization algorithm for discounted {0-1} knapsack problem,’ Neural Comput&Applic, vol. 30, no. 10, pp. 3019–3036.

Feng, Y, Wang, G-G, Li, W, Li, N 2018, ‘Multi-strategy monarch butterfly optimization algorithm for discounted {0-1} knapsack problem,’ NeuralComput&Applic, vol. 30,no. 10, pp. 3019–3036.

Gadekallu, TR, Khare, N, Bhattacharya, S, Singh, S, Maddikunta, PKR, Ra, IH &Alazab, M 2020, 'Early detection of diabetic retinopathy using PCA-firefly based deep learning model,' Electronics, vol. 9, no. 2, pp. 274.

Gao, H, Wei, X, Liao, J, Wang, R, Xu, J, Liu, X, Pan, X, Li, Z, Li, Z, Xia, Y & Wang, Q 2015, 'Lower bone mineral density in patients with Parkinson's disease: a cross-sectional study from Chinese Mainland,' Frontiers in aging neuroscience, vol. 7, no. 203, pp. 1-9.

Geron, A 2019, 'Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media, Newton, MA, USA,.

Gokalp O, Tasci E, Ugur A 2020, 'A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification,' Expert Syst Appl, vol. 146, pp. 1-30.

Gong, D, Xu, B, Zhang, Y, Guo Y & Yang S 2020, 'A Similarity-Based Cooperative Coevolutionary Algorithm for Dynamic Interval Multiobjective Optimization Problems,' IEEE Transactions on Evolutionary Computation, vol. 24, no. 1,pp. 142-156.

Greff, K, Srivastava, RK, Koutník, J, Steunebrink, BR &Schmidhuber, J 2016, 'LSTM: A search space odyssey,' IEEE transactions on neural networks and learning systems, vol. 28, no. 10, pp. 2222-2232.

Grover, S, Bhartia, S, Yadav, A &Seeja, KR 2018, 'Predicting severity of Parkinson's disease using deep learning,' Procedia computer science, vol. 132, pp. 1788-1794.

Guan, D, Yuan, W, Lee, Y, Najeebullah, K and Rasel, MK 2014, 'A review of ensemble learning based feature selection,' IETE Technical Review, vol. 31, no. 3, pp. 190-198.

Guerrero, M, Castillo, O & García, M 2015, 'Cuckoo search via lévy flights and a comparison with genetic algorithms,' In Fuzzy logic augmentation of nature-inspired optimization metaheuristics (pp. 91-103). Springer, Cham.

Guerrero-Luis, M, Valdez, F & Castillo, O 2021, ‘A Review on the Cuckoo Search Algorithm,’ In Fuzzy Logic Hybrid Extensions of Neural and Optimization Algorithms: Theory and Applications (pp. 113-124). Springer, Cham.

Gunduz, H 2019, ‘Deep learning-based Parkinson’s disease classification using vocal feature sets,’ IEEE Access, vol. 7, pp. 115540-115551.

Gupta, D, Julka, A, Jain, S, Aggarwal, T, Khanna, A, Arunkumar, N & de Albuquerque, VHC 2018, ‘Optimized cuttlefish algorithm for diagnosis of Parkinson’s disease,’ Cognitive systems research, vol. 52, pp. 36-48.

Gupta, D, Sundaram, S, Khanna, A, Hassanien, AE & De Albuquerque, VHC 2018, ‘Improved diagnosis of Parkinson's disease using optimized crow search algorithm,’ Computers & Electrical Engineering, vol. 68, pp. 412-424.

Halbgebauer, S, Nagl, M, Klafki, H, Haußmann, U, Steinacker, P, Oeckl, P, Kassubek, J, Pinkhardt, E, Ludolph, AC, Soininen, H & Herukka, SK 2016, ‘Modified serpinA1 as risk marker for Parkinson’s disease dementia: Analysis of baseline data,’ Scientific reports, vol. 6, no. 1, pp. 1-8.

Haq, AU, Li, J, Memon, MH, Khan, J, Din, SU, Ahad, I, Sun, R & Lai, Z 2018, ‘Comparative analysis of the classification performance of machine learning classifiers and deep neural network classifier for prediction of Parkinson disease,’ In 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) (pp. 101-106). IEEE.

Haq, AU, Li, JP, Memon, MH, Malik, A, Ahmad, T, Ali, A, Nazir, S, Ahad, I & Shahid, M 2019, ‘Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson’s disease using voice recordings,’ IEEE access, vol. 7, pp. 37718-37734.

Hasan, KA & Hasan, MAM 2020, ‘Classification of Parkinson's Disease by Analyzing Multiple Vocal Features Sets,’ In 2020 IEEE Region 10 Symposium (TENSYMP), pp. 758-761.

Hassan, MR, Kadir, SO, Islam, MA, Abujar, S &Zannat, R 2019, 'A Knowledge Base Data Mining based on Parkinson's Disease,' In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) (pp. 14-17).

Hosseini, FS 2020, 'Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection,' Sci Total Environ. Vol. 711, pp. 135161.

Hssayeni, MD, Jimenez-Shahed, J, Burack, MA &Ghoraani, B 2019, 'Symptom-based, dual-channel LSTM network for the estimation of unified Parkinson's disease rating scale iii,' In 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (pp. 1-4). IEEE.

Ibrahim, HT, Mazher, WJ, Ucan, ON & Bayat, O 2019, 'A grasshopper optimizer approach for feature selection and optimizing SVM parameters utilizing real biomedical data sets,' Neural Computing and Applications, vol. 31, no. 10, pp. 5965-5974.

Islam, MS, Parvez, I, Deng, H & Goswami, P 2014, 'Performance comparison of heterogeneous classifiers for detection of Parkinson's disease using voice disorder (dysphonia),' In 2014 International Conference on Informatics, Electronics & Vision (ICIEV), pp. 1-7.

Jain, D & Singh, V 2018, 'An efficient hybrid feature selection model for dimensionality reduction,' Procedia Computer Science, vol. 132, pp. 333-341.

Jankowski, N 2013, 'Fast heterogeneous boosting,' In 2013 IEEE Symposium on Computational Intelligence and Ensemble Learning (CIEL), pp. 1-8.

Joshi, D, Khajuria, A & Joshi, P 2017, 'An automatic non-invasive method for Parkinson's disease classification,' Computer methods and programs in biomedicine, vol. 145, pp.135-145.

K. H. Leung, S. P. Rowe, M. G. Pomper, and Y. Du, "A three-stage, deep learning, ensemble approach for prognosis in patients with Parkinson's disease," EJNMMI Research, vol. 11, no. 1, pp. 1–14, 2021.

Kadam, VJ & Jadhav, SM 2019, 'Feature ensemble learning based on sparse autoencoders for diagnosis of Parkinson's disease,' In Computing, communication and signal processing (pp. 567-581). Springer, Singapore.

Kadkhodaei, HR, Moghadam, AME &Dehghan, M 2020, 'HBoost: A heterogeneous ensemble classifier based on the Boosting method and entropy measurement,' Expert Systems with Applications, vol. 157, pp. 1-14.

Karan, B, Sahu, SS & Mahto, K 2020, 'Parkinson disease prediction using intrinsic mode function-based features from speech signal,' Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 249-264.

Kaya, D 2019, 'Optimization of SVM parameters with hybrid CS-PSO algorithms for Parkinson's disease in LabVIEW environment,' Parkinson's Disease, 2019.

Khoury, N, Attal, F, Amirat, Y, Oukhellou, L & Mohammed, S 2019, 'Data-driven based approach to aid Parkinson's disease diagnosis,' Sensors, vol. 19, no. 2, pp. 242.

Kubota KJ, Chen, JA & Little, MA 2016. 'Machine learning for largescale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures,' Movement disorders, vol. 31, no. 9, pp. 1314–1326.

Lahmiri, S, Dawson, DA & Shmuel, A 2018, 'Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures,' Biomedical engineering letters, vol. 8, no. 1, pp. 29-39.

Lertampaiporn, S, Thammarongtham, C, Nukoolkit, C, Kaewkamnerdpong, B & Ruengjitchat chawalya, M 2013, 'Heterogeneous ensemble approach with discriminative features and modified-SMOTE bagging for pre-miRNA classification,' Nucleic acids research, vol. 41, no. 1, pp. e21-e21.

Li, Q, Chen, H, Huang, H, Zhao, X, Cai, Z, Tong, C, Liu, W & Tian, X 2017, 'An enhanced grey wolf Optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis,' Computational and mathematical methods in medicine, 2017.

Lopez-Rincon, A, Martinez-Archundia, M, Martinez-Ruiz, GU, Schoenhuth, A &Tonda, A 2019, 'Automatic discovery of 100-miRNA signature for cancer classification using ensemble feature selection,' BMC bioinformatics, vol. 20, no. 1, pp. 1-17.

Ly, QT, Homolosine, AA, Gilat, M, Chai, R, Martens, KAE, Georgiades, M, Naik, GR, Tran, Y, Lewis, SJ & Nguyen, HT 2017, 'Detection of turning freeze in Parkinson's disease based on S-transform decomposition of EEG signals,' In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 3044-3047). IEEE.

Mareli, M & Twala, B 2018, 'An adaptive Cuckoo search algorithm for Optimization,' Applied computing and informatics, vol. 14, no. 2, pp. 107-115.

Masood, S, Wisal, K, Pal, O & Kumar, C 2021, 'An ensemble feature selection framework for early detection of Parkinson's disease based on feature correlation analysis,' Authored Preprints.

Massano, J & Bhatia, KP 2012, 'Clinical approach to Parkinson's disease: features, diagnosis, and principles of management,' Cold Spring Harbor perspectives in medicine, vol. 2, no. 6, pp. 1-16.

Mera-Gaona, M, López, DM, Vargas-Canas, R & Neumann, U 2021, 'Framework for the Ensemble of Feature Selection Methods,' Applied Sciences, vol. 11, no. 17, pp. 1-16.

Mesa, A, Castromayor, K, Garillos-Manliguez, C & Calag, V 2018, 'Cuckoo search via Levy flights applied to uncapacitated facility location problem,' Journal of Industrial Engineering International, vol. 14, no. 3, pp. 585-592.

Mohapatra, S, Patra, D & Satpathy, S 2014, 'An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images,' Neural Computing and Applications, vol. 24, no. 7, pp. 1887-1904.

Moradi, P & Rostami M 2015, 'Integration of graph clustering with ant colony optimization for feature selection,' Knowl Based Syst.vol. 84, pp. 144–61.

Mostafa, SA, Mustapha, A, Khaleefah, SH, Ahmad, MS & Mohammed, MA 2018, 'Evaluating the performance of three classification methods in diagnosis of Parkinson's disease,' In International Conference on Soft Computing and Data Mining (pp. 43-52). Springer, Cham.

Mostafa, SA, Mustapha, A, Mohammed, MA, Hamed, RI, Arunkumar, N, Abd Ghani, MK, Jaber, MM & Khaleefah, SH 2019, 'Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease,' Cognitive Systems Research, 54, pp.90-99.

Nilashi, M, Ibrahim, O, Ahmadi, H, Shahmoradi, L & Farahmand, M 2018, 'A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques,' Biocybernetics and Biomedical Engineering, vol. no. 38, no. 1, pp. 1-15.

Nilashi, M.; Abumalloh, R.A.; Yusuf, S.Y.M.; Thi, H.H.; Alsulami, M.; Abosaq, H.; Alyami, S.; Alghamdi, A. Early diagnosis of Parkinson's disease: A combined method using deep learning and neuro-fuzzy techniques. *Comput. Biol. Chem.* **2023**, *102*, 107788.

Oh, SL, Hagiwara, Y, Raghavendra, U, Yuvaraj, R, Arunkumar, N, Murugappan, M & Acharya, UR 2020, 'A deep learning approach for Parkinson's disease diagnosis from EEG signals,' Neural Computing and Applications, vol. 32, no. 15, pp. 10927-10933.

Oliveira, FHM, Machado, AR & Andrade, AO 2018, 'On the use of t-distributed stochastic neighbor embedding for data visualization and classification of individuals with Parkinson's disease,' Computational and mathematical methods in medicine, 2018.

Oung, QW, Muthusamy, H, Basah, SN, Lee, H and Vijean, V 2018, 'Empirical wavelet transform based features for classification of Parkinson's disease severity,' Journal of medical systems, vol. 42, no. 2, pp. 1-17.

Parisi, L, Ravi Chandran, N & Manaog, ML 2018, 'Feature-driven machine learning to improve early diagnosis of Parkinson's disease,' Expert Systems with Applications, vol. 110, pp. 182-190.

Peker, M 2016, ‘A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM,’ Journal of medical systems, vol. 40, no. 5, pp. 1-16.

Peng, J, Guan, J & Shang, X 2019, ‘Predicting Parkinson's disease genes based on node2vec and autoencoder,’ Frontiers in genetics, vol. 10, pp. 226.

Pereira, CR, Weber, SA, Hook, C, Rosa, GH & Papa, JP 2016, ‘Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics,’ in 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) , pp. 340-346.

Pérez, CJ, Naranjo, L, Martín, J & Campos-Roca, Y 2014, ‘A latent variable-based Bayesian regression to address recording replications in Parkinson's Disease,’ In 2014 22nd European Signal Processing Conference (EUSIPCO) , pp. 1447-1451.

Pes, B 2020, ‘Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains,’ Neural Computing and Applications, vol. 32, no. 10, pp. 5951-5973.

Quan, C.; Ren, K.; Luo, Z. A deep learning based method for Parkinson’s disease detection using dynamic features of speech. *IEEE Access* **2021**, *9*, 10239–10252.

Ramezani, H, Khaki, H, Erzin, E & Akan, OB 2017, ‘Speech features for telemonitoring of Parkinson's disease symptoms,’ In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 3801-3805). IEEE.

Rasheed, J, Hameed, AA, Ajlouni, N, Jamil, A, Özyavaş, A & Orman, Z 2020, ‘Application of adaptive back-propagation neural networks for Parkinson’s disease prediction,’ In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI) (pp. 1-5). IEEE.

Reddy, GT, Reddy, MPK, Lakshmanna, K, Kaluri, R, Rajput, DS, Srivastava, G & Baker, T 2020, ‘Analysis of dimensionality reduction techniques on big data,’ IEEE Access, vol. 8, pp. 54776-54788.

Rizvi, DR, Nissar, I, Masood, S, Ahmed, M & Ahmad, F 2020, 'An LSTM based Deep learning model for voice-based detection of Parkinson's disease,' Int. J. Adv. Sci. Technol, vol. 29, no. 8, pp. 337-343.

Rokach, L 2009, 'Taxonomy for characterizing ensemble methods in classification tasks: a review and annotated bibliography,' Computational Statistics & Data Analysis, vol. 53, no. 12, pp. 4046–4072.

Rostami, M, Forouzandeh, S, Berahmand, K & Soltani, M 2020, 'Integration of multi-objective PSO based feature selection and node centrality for medical datasets,' Genomics, vol. 112, no. 6, pp. 4370-4384.

Rusz, J, Cmejla, R, Ruzickova, H & Ruzicka, E 2011, 'Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease,' The journal of the Acoustical Society of America, vol. 129, no. 1, pp. 350-367.

S. Xingjian, Z. Chen, H. Wang, D. Yeung, W. Wong, W. Woo 2015, 'Convolution LSTM network: A machine learning approach for precipitation nowcasting,' In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, pp. 802–810.

Sadeghian, Z, Akbari, E & Nematzadeh, H 2021, 'A hybrid feature selection method based on information theory and binary butterfly optimization algorithm,' Engineering Applications of Artificial Intelligence, vol. 97, pp. 104079.

Sakar CO, Serbes, G, Gunduz, A, Tunc, HC, Nizam, H, Sakar, BE, Tutuncu, M, Aydin, T, Isenkul, M & Apaydin, H 2019, 'A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable q-factor wavelet transform,' Applied Soft Computing, vol. 74, pp. 255–263.

Sakar, BE, Isenkul, ME, Sakar, CO, Sertbas, A, Gurgen, F, Delil, S & Kursun, O 2013, 'Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings,' IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 828-834.

Sakar, BE, Serbes, G & Sakar, CO 2017, 'Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease,' PloS one, vol. 12, no. 8, pp. 1-18,

Salmanpour, MR, Shamsi, M & Rahmim, A 2021, 'Feature selection and machine learning methods for optimal identification and prediction of subtypes in Parkinson's disease, 'Computer Methods and Programs in Biomedicine, vol. 206, pp. 106131.

Segovia, F, Górriz, JM, Ramírez, J, Martínez-Murcia, FJ & García-Pérez, M 2018, 'Using deep neural networks along with dimensionality reduction techniques to assist the diagnosis of neurodegenerative disorders,' Logic Journal of the IGPL, vol. 26, no. 6, pp. 618-628.

Sehgal, S, Agarwal, M, Gupta, D, Sundaram, S & Bashambu, A 2020, 'Optimized grass hopper algorithm for diagnosis of Parkinson's disease,' SN Applied Sciences, vol. 2, no. 6, pp. 1-18.

Seijo-Pardo, B, Bolón-Canedo, V, Porto-Díaz, I and Alonso-Betanzos, A 2015, 'Ensemble feature selection for rankings of features,' In International Work-Conference on Artificial Neural Networks ,pp. 29-42, Springer, Cham.

Seijo-Pardo, B, Porto-Díaz, I, Bolón-Canedo, V & Alonso-Betanzos, A 2017, 'Ensemble feature selection: homogeneous and heterogeneous approaches,' Knowledge-Based Systems, vol. 118, pp. 124-139.

Selesnick IW 2011, 'Wavelet transform with tunable q-factor,' IEEE transactions on signal processing, vol. 59, no. 8, pp. 3560–3575,

Semwal, VB, Singha, J, Sharma, PK, Chauhan, A & Behera, B 2017, 'An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification,' Multimedia tools and applications, vol. 76, no. 22, pp. 24457-24475.

Sarankumar R., Vinod D., Anitha K., Manohar G., Vijayanand K.S., Pant B., Sundramurthy V.P. 2022, 'Severity Prediction over Parkinson's Disease Prediction by Using the Deep Brooke Inception Net Classifier'. Computational Intelligent. Neuroscience:7223197. doi: 10.1155/2022/7223197

Shahbakhi, M, Far, DT & Tahami, E 2014. ‘Speech analysis for diagnosis of parkinson’s disease using genetic algorithm and support vector machine,’ Journal of Biomedical Science and Engineering, vol. 7, no. 4, pp. 147–156.

Shao, R, Hu, W, Wang, Y & Qi, X 2014, ‘The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet, Measurement, vol. 54, pp. 118-132.

Shapira, AL, Handzel, R & Korczyn, AD 2017, ‘The lived experience of parkinson's ‘disease: A content analysis of parkinson's patients' blogs,’ Isr Med Assoc J, vol. 19, no. 11, pp. 685-690.

Sharma, A & Giri, RN 2014, ‘Automatic recognition of Parkinson’s Disease via artificial neural network and support vector machine,’ International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 4, no. 3, pp. 2278-3075.

Sharma, P, Sundaram, S, Sharma, M, Sharma, A & Gupta, D 2019, ‘Diagnosis of Parkinson’s disease using modified grey wolf optimization,’ Cognitive Systems Research, vol. 54, pp. 100-115.

Shehab, M, Khader, AT & Al-Betar, MA 2017, ‘A survey on applications and variants of the cuckoo search algorithm,’ Applied Soft Computing, vol. 61, pp. 1041-1059.

Shi, X, Chen, Z, Wang, H, Yeung, DY, Wong, WK and Woo, WC 2015, ‘Convolutional LSTM network: A machine learning approach for precipitation nowcasting,’ Advances in neural information processing systems, vol. 28, pp. 802–810.

Shrivastava, P, Shukla, A, Vepakomma, P, Bhansali, N & Verma, K 2017, ‘A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease,’ Computer methods and programs in biomedicine, vol. 139, pp. 171-179.

Sim, J, Park, JS, Kim, M, Bae, D, Choi, Y & Kim, LS 2016, ‘14.6 a 1.42 tops/w deep convolutional neural network recognition processor for intelligent ioe systems,’ In 2016 IEEE International Solid-State Circuits Conference (ISSCC) (pp. 264-265). IEEE.

Singh, N & Singh, P 2021, 'A hybrid ensemble-filter wrapper feature selection approach for medical data classification,' Chemometrics and Intelligent Laboratory Systems, vol. 217, pp. 104396.

Solana-Lavalle, G, Galán-Hernández, JC & Rosas-Romero, R 2020, 'Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features,' Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 505-516.

Srinivasan, SM, Martin, M & Tripathi, A 2017, 'ANN based Data Mining Analysis of the Parkinson's Disease,' International Journal of Computer Applications, vol. 168, no. 1, pp. 56-60.

Sriram, TV, Rao, MV, Narayana, GS & Kaladhar, DSVGK 2015, 'Diagnosis of Parkinson Disease Using Machine Learning and Data Mining Systems from Voice Dataset,' In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA), pp. 151-157.

Sultana, N & Islam, MM 2020, 'Meta classifier-based ensemble learning for sentiment classification',in Proceedings of International Joint Conference on Computational Intelligence (pp. 73-84). Springer, Singapore.

Tang, J, Alelyani, S & Liu, H 2014, 'Feature selection for classification: A review,' Data classification: Algorithms and applications, pp. 1-33.

Thomas, P, El Haouzi, HB, Suhner, MC, Thomas, A, Zimmermann, E, Noyel, M 2018, 'Using a classifier ensemble for proactive quality monitoring and control: the impact of the choice of classifiers types, selection criterion, and fusion process,' Computers in Industry, vol. 99, pp. 193–204,

Tsanas, A, Little, MA, McSharry, PE, Spielman, J & Ramig, L O 2012, 'Novel speech signal processing algorithms for highaccuracy classification of Parkinsons disease,' IEEE Transactions on Biomedical Engineering, vol. 59, no. 5, pp. 1264–1271.

Tubishat, M, Idris, N, Shuib, L, Abushariah, MA & Mirjalili, S 2020, 'Improved salp swarm algorithm based on opposition based learning and novel local search algorithm for feature selection,' Expert Systems with Applications, vol. 145, pp. 113122.

Tuncer, T, Dogan, S & Acharya, UR 2020, 'Automated detection of Parkinson's disease using minimum average maximum tree and singular value decomposition method with vowels,' Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 211-220.

T. E. Aravindan1&R. Seshasayanan, 2018,'Denoising Brain Images with the Aid of Discrete Wavelet Transformand Monarch Butterfly Optimization with Different Noises,'J ournal of Medical Systems, vol. 42, pp. 207.

van Rijn, JN, Holmes, G, Pfahringer, B & Vanschoren, J 2018, 'The online performance estimation framework: heterogeneous ensemble learning for data streams,' Machine Learning, vol. 107, no. 1, pp. 149-176.

Wahid, F, Begg, RK, Hass, CJ, Halgamuge, S & Ackland, DC 2015, 'Classification of Parkinson's disease gait using spatial-temporal gait features,' IEEE journal of biomedical and health informatics, vol. 19, no. 6, pp. 1794-1802.

Wang, GG, Deb, S & Cui, Z 2019, 'Monarch butterfly optimization,' Neural computing and applications, vol. 31, no. 7, pp. 1995-2014.

Wang, GG, Zhao, X & Deb, S 2015, 'A novel monarch butterfly optimization with greedy strategy and self-adaptive,' In Second International Conference on Soft Computing and Machine Intelligence (ISCMI) ,pp. 45-50.

Wang, H, Wang, W, Zhou, X, Sun, H, Zhao, J, Yu, X & Cui, Z 2017 'Firefly algorithm with neighborhood attraction,' Information Sciences, vol. 382, pp. 374-387.

Wang, Q 2012, 'Kernel principal component analysis and its applications in face recognition and active shape models,' arXiv preprint arXiv:1207.3538.

Wang, SH, Zhan, TM, Chen, Y, Zhang, Y, Yang, M, Lu, HM, Wang, HN, Liu, B & Phillips, P 2016, 'Multiple sclerosis detection based on biorthogonal wavelet transform,' RBF kernel principal component analysis, and logistic regression IEEE Access, vol. 4, pp. 7567-7576.

Wang, W, Wang, H, Zhao, J & Lv, L 2017, 'A new adaptive firefly algorithm for solving optimization problems,' In International Conference on Intelligent Computing (pp. 649-657). Springer, Cham.

Welikala, RA, Fraz, MM, Dehmeshki, J, Hoppe, A, Tah, V, Mann, S, Williamson TH & Barman SA 2015, 'Genetic algorithm-based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy', Computerized Medical Imaging and Graphics, no. 43, pp. 64-77.

Xingjian, S, Chen, Z, Wang, H, Yeung, D, Wong, W & Woo, W 2015, 'Convolutional LSTM network: A machine learning approach for precipitation nowcasting,' in Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, vol. 2015, pp. 802–810

Xingjian, S, Chen, Z, Wang, H, Yeung, D,Wong, W, Woo, W 2015 ,'Convolutional LSTM network: A machine learning approach for precipitation nowcasting,' In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, ; pp. 802–810.

Xue, B, Zhang, M, Browne, WN & Yao, X 2015, 'A survey on evolutionary computation approaches to feature selection,' IEEE Transactions on Evolutionary Computation, vol. 20, no. 4, pp. 606-626.

Yan, C, Ma, J, Luo, H & Patel, A 2019, 'Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets,' Chemometrics and Intelligent Laboratory Systems, vol. 184, pp. 102-111.

Yang, XS & Deb, S 2014, 'Cuckoo search: recent advances and applications. Neural Computing and applications, vol. 24, no. 1, pp. 169-174.

Yang, XS & He, X 2013, 'Firefly algorithm: recent advances and applications,' International journal of swarm intelligence, vol. 1, no. 1, pp. 36-50.

Yang, Z, Gjorgjevikj, D, Long, J, Zi, Y, Zhang, S & Li, C 2021, 'Sparse autoencoder-based multi-head deep neural networks for machinery fault diagnostics with detection of novelties,' Chinese Journal of Mechanical Engineering, vol. 34, no. 1, pp. 1-12.

Yong, Z, Dun-wei G & Wan-qiu Z 2016, 'Feature selection of unreliable data using an improved multi-objective PSO algorithm,' Neurocomputing, vol. 171, pp. 1281-1290.

Yunusova Y, Weismer, G, Westbury, JR & Lindstrom, MJ 2008. 'Articulatory movements during vowels in speakers with dysarthria and healthy controls,' Journal of Speech, Language, and Hearing Research,

Zham, P, Arjunan, SP, Raghav, S & Kumar, DK 2017, 'Efficacy of guided spiral drawing in the classification of Parkinson's disease,' IEEE journal of biomedical and health informatics, vol. 22, no. 5, pp. 1648-1652.

Zhao, S, Rudzicz, F, Carvalho, LG, Márquez-Chin, C & Livingstone, S 2014, 'Automatic detection of expressed emotion in Parkinson's disease,' In 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4813-4817). IEEE.

LIST OF PUBLICATIONS

1. B. Sabeena Dr.S. Sivakumari Implementation of Feature Selection and Ensemble Deep Learning Classifier for Parkinson's disease "Oxidation Communications" Volume no: 45 No.4, pp.701-713 December 2022 (Scopus indexed).

2. B. Sabeena Dr.S. Sivakumari Ensemble Feature Selection and Ensemble Deep Learning Classifier for Parkinson's disease "Journal of Theoretical and Applied Information Technology "Vol.101.no, pp.3416-3430 May 2023 (Scopus indexed).

Avinashilingam Institute for HomeScience and Higher Education for Women

(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category 'A' by MHRD
Re-accredited with A++ Grade by NAAC. CGPA 3.65/4, Category I by UGC
Coimbatore - 641 043, Tamil Nadu, India

Appendix L2

(Item No 5 of Check List)

Details of Research Publications

S.No	Article	Journal	Other Details Vol/No/Page No/ Year	Published in UGC-CARE / Scopus Indexed/ Web of Science
1	Implementation Of Feature Selection And Ensemble Deep Learning Classifier For Parkinson's Disease	Oxidation Communications	Volume 45, No 4, pp.701–713 December 2022	Scopus Indexed The journal's topics of interest includes medical image processing, Neural networks & fuzzy logic techniques
2	Ensemble Feature Selection and Ensemble Deep Learning (Edl) Classifier for Parkinson's Disease".	Journal of Theoretical and Applied Information Technology	Vol.101. No 9 pp.3416-3430 May 2023	Scopus Indexed

*Proof of list of Journals from Internet to be attached along with copies of reprints.

Scholar : [signature]

Supervisor : [signature] 10/8/2023

Checked By: [signature] 11/8/23

[signature] 10/8/2023

HoD/Dean of Respective School

2

The scholar Miss. Sabeena, B (19PHEOF002) has published her paper in the following journals:

1. Journal of Theoretical and Applied Information Technology - indexed and active in Scopus from 2008 to present. The scholar published her article in May 2023 issue (Vol. 101, No. 9, May, Pg. 3416-3430, 2023) and
2. Oxidation Communications - indexed and active in Scopus from 1996 to present. The scholar published her paper in December 2022 (Vol. 45, No. 4, Pg. 701-713, December, 2022).

[signature]
08.08.23

Kernel based Principal Component Analysis based Dimensionality Reduction and Fuzzy Convolution Long Short-term Memory based Convolutional Neural Network (FCLSTM-CNN) for Parkinson's Disease Classification

B. Sabeena, Research Scholar, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering. E-mail: sabeena.bheeman@gmail.com

S. Sivakumari, Professor, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering. E-mail: prof.sivakumari@gmail.com

P. Amudha, Associate Professor, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering. E-mail: amudharul@gmail.com

Abstract--- A type of progressive neurodegenerative disease is Parkinson's Disease (PD) with multiple motor and non-motor characteristics. In early stage of this disease, vocal impairments is commonly faced by PD patients. So, on recent PD detection studies, vocal disorders based diagnosis system is focused a lot and in data mining, there will be a use of feature pace's dimension reduction. In various studies, investigated about, optimal number of dimensions reduction with the aim of computing number of non-trivial components. So, for solving this issue, from dataset, dimensions are reduced by introducing Kernel based Principal Component Analysis (KPCA). For evaluating dimensionality reduction effects, classification is performed in two stages. In first stage, an algorithm is developed for working in original feature space and compute reduced feature subset and classifier is applied. In second stage, a new algorithm is introduced for using new feature space properties directly via dimensionality reduction and transformation. With various features combination like Vocal fold features, Tunable Q-factor Wavelet Transform (TQWT),Wavelet Transform (WT),Mel-Frequency Cepstral Coefficients (MFCC),Time frequency features, employed Finally Fuzzy Convolution Long Short-Term Memory based Convolutional Neural Network (FCLSTM-CNN) classifier. There will be difference in feature set combination. Fuzzy membership functions are used for computing bias and weight values in proposed classifier. Dataset taken from University of California-Irvine (UCI) Machine Learning repository is used for training proposed FCLSTM -CNN classifier. Leave-One-Person-Out Cross Validation (LOPO CV) is used for validating performance of proposed technique. Classification metrics are used for measuring the results of these classifiers. Along with accuracy, Matthews Correlation Coefficient (MCC) and F-measure are used for assessment because of data classes imbalance distribution.

Keywords--- Vocal Features, Fuzzy Convolution Long Short-Term Memory based Convolutional Neural Network (FCLSTM - CNN), Deep Learning, Dimensionality Reduction, Kernel based Principal Component Analysis (KPCA), Health Informatics, Parkinson's Disease (PD) Classification.

I. Introduction

A type of chronic neurodegenerative disorder of nervous system is Parkinson's Disease (PD). Motor functions are affected by this predominantly. It is a movement disorder class, which includes shaking movement in resting position (Parkinson's tremor), increased muscle tonus (rigidity), diminished and slow movement (bradykinesis), inability of voluntary movement (akinesis) features [1]. Other features like balance problem, voice and speech characteristics changes and diminished facial expressions are included in it [2,3].

Sense of smell (anosmia) will be lost by people having PD and during rapid eye movement sleep (REMs) phase, they tend to have sleep disorders [4]. Around 1% of population will be affected by PD as estimated [5]. PD patient's detection needs a highly accurate as well as reliable health informatics systems because of patients prolonged life with an early diagnosis. Civilians workload can also be reduced using this system [6 – 8].

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

In an early diagnosis, for enhancing objectivity and assistance, medical decision support tools are used. For patients, specific treatments design are allowed by this early diagnosis [9]. Accurate biomarkers identification are the major goal of neurodegenerative diseases [10]. For detection of PD, there is a huge amount of studies in literature which focuses on speech processing [11-12].

Natural speech and sustained vowels are used for diagnosis. For predictions, expert support tools are offered by classification algorithms and other intelligent techniques. Clinically relevant features can be obtained using various speech signal processing techniques and for obtaining reliable decisions in PD classification, these features are fed to different artificial learning techniques.

In classification of PD, commonly used algorithms includes L1-Norm Support Vector Machine (SVM), Kernel-Based Extreme Learning Machine (KELM), Artificial Neural Networks (ANN) and because of easy understanding and simplicity, K-Nearest Neighbors (KNN) and Random Forest (RF) methods are also used.

From the data, selected features quality directly defines the mentioned algorithms success. Also, manual selection of relevant features are highly difficult as speech data's intrinsic properties are represented by this relevant features and deep learning methods can also be used for automatic learning of latent properties.

Kotsavasiloglou et al [13] differentiated pathological and normal movement in PD by establishing a simple as well as objective metrics. In differentiating Parkinson's disease and healthy individuals, better performance can be exhibited using proposed technique and it has promising applications in telemedicine applications.

Gallicchio et al [14] implemented a deep Echo State Networks (ESNs) based PD diagnosis technique. Entire time series collected using a tablet device is analysed for performing PD identification during spiral tests sketching. It doesnot require any feature extraction and data pre-processing.

Sakar et al [8]implemented a two-step PD detection technique. In first step, only patient data is used for computing patient group having high speech impairments severity via Unified Parkinson's Disease Rating Scale (UPDRS) score as a disease progression index. In next step, this group of patients samples are excluded from dataset and new dataset with PD patients and healthy subjects samples with less speech impairments severity is created. Three classifiers having different settings for addressing binary classification problem.

Khatamino et al [15] implemented a Deep Learning system based on Convolutional Neural Network (CNN) for learning features of Handwriting drawing spirals. These drawings are made by People with Parkinson.

Gil-Martín et al [16] implemented a PD detection technique based on Convolutional Neural Network (CNN) for detecting PD from drawing movements. There are two parts in this CNN. They are convolutional layer based feature extraction and fully connected layer based classification. Fast Fourier's transform modules are given as an input to CNN with the frequency range of 0 to 25 Hz. Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet dataset is used performing the analysis.

In data, non-complex and complex relationships can be modelled using Deep Neural Network (DNN) with its ability. PD classification can also be done using this. In PD classification studies, classification methods based on deep learning is introduced still now. But, for improving PD classification results, dimensionality reduction is also performed in this study.

Classifier's running time is reduced by minimizing the features or samples dimension and PD classification accuracy can also be increased using this technique. Set of speech or vocal features are used for PD classification in this work and it proposed a deep FCLSTM-CNN classifier for the same [15]. A CNN based framework is build for this purpose, which assist us in distinguishing PD patients and healthy individuals. For PD detection classification, extracted some features from input dataset.

Kernel Principal Component Analysis (KPCA) is used for reducing the dimensionality of extracted features. The minimum Redundancy Maximum Relevance (mRMR) is used for selecting informative features from entire features set. According to class label, features having high relevance score are selected by mRMR and redundant features are eliminated using this. According to the relevance between target and PD features, PD features are ranked by mRMR technique. It also penalizes PD features redundancy.

At last, correlation between genes are explored using FCLSTM classifier and genes are predicted one by one. Dataset derived from University of California-Irvine (UCI) Machine Learning repository is used for training this classifier. Three voice records per individual are included in that dataset. When compared to SCFW-KELM, SVM, CNN classifiers, highly accurate results are produced by proposed FCLSTM-CNN classifier as shown in results.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

II. Literature Review

Recent studies on PD classification are summarized in this section. Techniques using machine learning algorithms are also summarized and also recent learning techniques are covered.

Machine Learning Methods

Berus et al [17] implemented multiple feed-forward Artificial Neural Networks (ANNs) having different configuration are utilized for predicting PD individuals according to extracted features from 26 various voice samples per individuals. At last, fine-tune NN and around 86.47% accuracy is achieved. Other feature selection procedures can be used for enhancing performance of ANN.

Haq et al [18] predicted PD by implementing L1-norm Support Vector Machine (SVM) model. Highly appropriate as well as related features are selected using L1-norm SVM based feature selection technique, which results in accurate healthy and PDD people classification. But, this work does not involve any dimensionality reduction before classification, so, it enhances the PD classifier's computation time.

Tsanas et al [19] discriminated healthy people from PD subjects by implementing novel algorithms like Support Vector Machines (SVMs) and Random Forests (RFs). From sustained vowels, 132 dysphonia measures are computed. Four algorithms of feature selection are used for selecting these dysphonia measures four parsimonious subsets and these feature subsets are mapped to binary classification. They concluded that, from Relief selection, feature set having low classification error can be obtained.

Rouzbahani and Daliri [20] used voice signals for suggesting PD detection model. Parameters like HNR, pitch, shimmer, jitter, fundamental frequency are used for defining inputs of this proposed model along with statistical measures based on these parameters.

Various feature selection techniques like ROC curves, t-test, Fisher's Discriminant Ratio and correlation rates are used for selecting informative features from these entire feature set. Training of Discrimination-Function, KNN and SVM based classifiers are done after computing optimum features. Metrics like specificity, sensitivity, error rate, accuracy are used for measuring classifier performances and KNN classifier produces better performance with around 93.82% of accurate results.

Parisi et al [21]aided early PD diagnosis by applying a novel hybrid Artificial Intelligence-based classifier. Features are selected using weights which are computed using Multi-Layer Perceptron (MLP) and based on its relative importance in discriminating pathological and physiological data patterns, input features are ranked using their moduli.

Commercially available classifiers and software based studies are used for comparing performance of this hybrid feature-driven algorithm (MLP-Lagrangian Support Vector Machine (LSVM)). With relatively fast convergence , enhancement in overall accuracy in classification is achieved by proposed method as indicated in results and it demonstrates its potential in PD's early diagnosis.

Deep Learning Methods

Wroge et al [22] used supervised classification algorithms like deep neural networks for explotring its effectiveness in accurate disease diagnosis of individuals. With pathological post-mortem examination as ground truth, around 85% of accurate results are given by machine learning models, which is high when compared with 73.8% average clinical diagnosis accuracy of non-experts, 79.6%average accuracy of movement disorder specialists without follow-up and 83.9% with follow-up.

Gürüler [23]combined Complex-Valued Artificial Neural Network (CVANN) and K-Means Clustering-based Feature Weighting (KMCFW). Also, application confirms conclusion that reliability of complex-valued algorithm's classification ability with regard to a real-valued dataset is high.

Ma et al [24] integrated effective subtractive clustering features weighting for proposing a hybrid technique called subtractive clustering features weighting Kernel-Based Extreme Learning Machine (SCFW-KELM) and for PD diagnosis, they introduced an fast classifier KELM. Because of its excellent performance, proposed technique can be utilized as a new powerful candidate for PD diagnosis.

Pereira et al [25] used CNN for proposing PD automatic identification technique. From an extracted signal, this method learns features during individual's exam using a smart pen, which is composed of sensor series.

From handwritten dynamics information can be extracted using this sensors. Also, they have created a public dataset and it is used for foster PD-related research by researchers in worldwide.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

Eskofier et al [26] implemented a CNN based deed learning model. Other state-of-the-art machine learning algorithms are outperformed by this deep learning model as indicated in results with around 4.6% enhancement in classification rate. For classification, it does not involve CNN having various parallel layers.

Gunduz [27] used vocal (speech) features set for classifying PD by implementing two Convolutional Neural Networks based frameworks.

In first framework, before given to 9-layered CNN, various features are combined. In second framework, parallel input layers are given with features and this layers are having direct connection with convolutional layers.

Before combining in merge layer, simultaneously extract deep features from every parallel branch. Before the task of classification, this work does not involve dimensionality reduction, which enhances the PD classifier's computation time.

Sakar et al [28] proposed a tunable Q-factor wavelet change (TQWT) to the voice signs of PD patients for highlight extraction, which has higher recurrence goals than the old style discrete wavelet change.

The component subsets are taken care of to numerous classifiers and the expectations of the classifiers are joined with troupe learning draws near. As opposed to utilizing separate element types in model preparing, most examinations utilize the blend of individual component types to perform arrangement task.

Expanded component space in these examinations can be decreased by means of highlight choice strategies [28]. In spite of the fact that, there are loads of side effects among the individuals exposing to the PD including eased back development, stance and equalization inadequacies, dysphonia which is characterized as the adjustments in discourse and enunciation, is the most important trailblazer of PD. This is the motivation behind why numerous examinations are centred around discourse based PD order.

PD patients predominantly face vocal rebellions which legitimately impact the vocal uproar, insecurity and recurrence variation from the norm.

previously mentioned investigations are analyzed, unmistakably related PD concentrates by and large use voice-based highlights with machine-based learning calculations. In spite of the fact that these investigations have utilize vocal-based highlights to manage PD characterization.

III. Proposed Methodology

Before giving specified dataset into PD classifications, its dimensionality is reduced using Kernel based Principal Component Analysis (KPCA) dimensionality algorithm in this work. Minimum Redundancy Maximum Relevance (mRMR) technique is employed for selecting informative features. Features having high relevance score based on class label are selected by mRMR and redundant features are eliminated by it. There are 60 features in that feature set, which are formed using MFCC.

Proposed Fuzzy Convolution Long Short-Term Memory based Convolutional Neural Network (FCLSTM-CNN). FCLSTM-CNN classifier is given with both feature sets for combining various vocal features types at feature and model level for distinguishing healthy individuals and PD patients. Proposed work's performance is evaluated using Leave-One-Person-Out Cross Validation (LOPO CV) procedure and proposed algorithm's performance is compared using SVM as a benchmark model. Figure 1 shows the proposed system's overall flowchart.

Parkinson's Disease Dataset

From UCI Machine learning repository, data is collected for using in this study and in study of [28], it is used recently.

In Cerrahpasa Faculty of Medicine, Istanbul University, at Neurology Department, gathered this dataset and it has 64 healthy individuals of 41 women and 23 men and 188 PD patientsof 81 women and 107 men.

Age of healthy patients lies between 41 to 82 years, while age of PD patients lies between33 to 87. Microphone's frequency response was set at 44.1 KHz during data collection and after doctor's review, vowel /a/ letter's repeated repetition in every person is collected with three replicates.

Feature Extraction

From dataset, extracted the features like Vocal fold, Tunable Q-factor Wavelet Transform (TQWT),Wavelet Transform,Mel-Frequency Cepstral Coefficients (MFCC), Time frequency Features and Baseline features.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

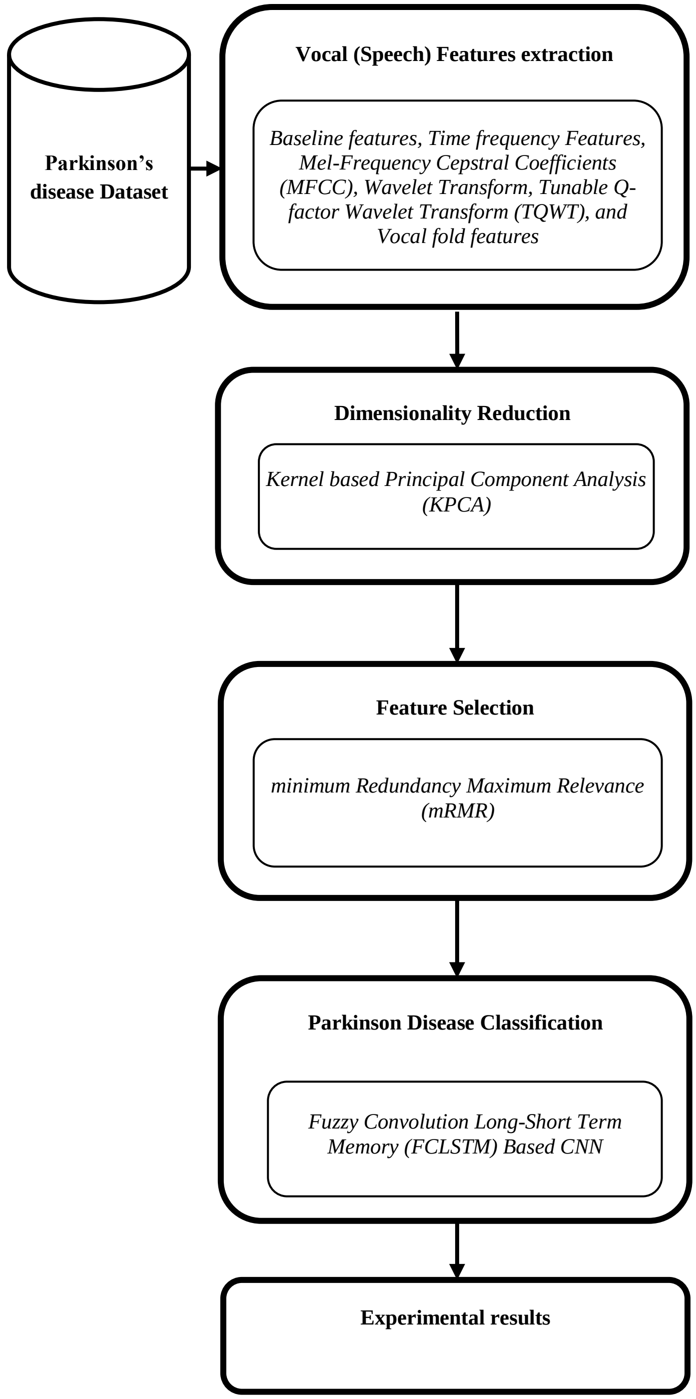


Figure 1: Overall Flow Chart of the Proposed System

i. Baseline Features

Even in early period, speech can be affected by PD as shown in literature [29] and so, PD can be evaluated using speech characteristics successfully. After medical treatment, it can be used for monitoring its evolution.

In studies [7], [19], commonly used characteristics of speech includes, Pitch Period Entropy (PPE), Detrended Fluctuation Analysis (DFA), Recurrence Time Density Entropy (RPDE),harmonicity parameters,fundamental frequency parameters and Jitter and glow based features. These characteristics are termed as baseline features in obtained data [28]. Description (#5) represents that there are 5 features are extracted from jitter and this common for all features.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X
Received: 04 May 2020/Accepted: 05 June 2020

Jitter: In fundamental frequency, it detects cycle-to-cycle changes (#5).

Shimmer variants: In fundamental amplitude, it detects cycle-to-cycle changes (#6).

Fundamental frequency parameters: Vocal fold vibration's maximum and minimum frequency values, standard deviation, median and mean (#5).

Harmonicity parameters: It quantifies signal information ratio over noise (in PD speech samples, increased noise components occurs) (#2).

Recurrence Period Density Entropy (RPDE): Vocal folds ability for providing stable vocal fold oscillations (#1)

Detrended Fluctuation Analysis (DFA): It quantifies turbulent noise's stochastic self-similarity (#1).

Pitch Period Entropy (PPE): For measuring fundamental frequency's impaired controls of via logarithmic scale(#1).

ii. Time Frequency Features

Bandwidth, Formant Frequencies and Intensity Parameters features are included in this.

Intensity Parameters: The speech signal power in dB (maximum, minimum and mean intensity) (#3).

Formant Frequencies: Frequencies which are amplified by vocal tract (first four formants) (#4).

Bandwidth: Frequency range between formant frequencies (first four bandwidths) (#4).

iii. Mel-Frequency Cepstral Coefficients (MFCCs)

Spectral domain partitioning and cepstral analysis are combined in MFCCs extraction technique using triangular overlapping filter banks. In articulators moment like lips and tongue, rapid deteriorations are detected using MFCCs in PD studies, which are affected directly by PD [7].

In data, there exist around 84 MFCCs related features and this features are derived from original 13 MFCCs mean and standard deviation in addition to signal's long energy and with their first and second derivatives [28]. In vocal tract, for catching PD affects separately from vocal folds (#84).

iv. Wavelet Transform (WT)

In a signals with small fluctuations in regional scale, decisions are made using a prominent tool called Wavelet Transform (WT). Various studies applies Particular features computed via WT from raw basic speech signal frequency (F_0), for diagnosing PD.

To extract WT-based features derived from raw (F_0) contour and log transformation of (F_0) contour, speech signal is applied with 10-level discrete wavelet transformation in collected data. Around 182 features based on WT which includes Teager-Kaiser energy of detailed and approximation coefficients, Shannon's and log energy entropy energy are produced by this process.

v. Tunable Q-Factor Wavelet Transform (TQWT)

Features can be extracted using another technique called Tunable Q-factor Wavelet Transform (TQWT). Advantages of three tuneable parameters namely Q- Q factor, r-redundancy and J-levels count are includes in TQWT for transforming signals in better quality based on signal's behaviour.

In time domain, in signals, oscillations count are related to Q-factor parameter. In decomposition stage, number of levels are represented as J. After decomposition, there are J+1 sub-bands produced by one final low-pass filter and J high pass filter outputs. In time, wavelet can be localized by parameter r, which controls excessive ringing without any affect in its shape [30]. Around 432 TQWT-related features are produced in this dataset with various experimentation [28].

vi. Vocal Fold Features

On vocal fold, for exploring noise effects, employed vocal fold vibration based features. For the same, features like, and Empirical Mode Decomposition (EMD), Vocal Fold Excitation Ratio (VFER), Glottal to Noise Excitation (GNE) and Glottis Quotient (GQ) are utilized [28].

Glottis Quotient (GQ): It specifies glottis opening and closing durations (#3).

Glottal to Noise Excitation (GNE): It quantifies turbulent noise's limit.Incomplete vocal fold closure produces this (#6).

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X
Received: 04 May 2020/Accepted: 05 June 2020

Vocal Fold Excitation Ratio (VFER): It quantifies the noise quantity produced because of pathological vocal fold vibration (#7).

Empirical Mode Decomposition (EMD): Adaptive basis functions are used for forming elementary signal components by decomposing speech signal and from these components, entropy and energy values are computed.

vii. Concat Features

Time frequency, vocal fold and baseline features combination forms Concat features.

Dimensionality Reduction by Kernel Principal Component Analysis (KPCA)

Dimension can be reduced using a popular method called Kernel Principal Component Analysis (KPCA).

Lowe dimensions linear subspace is computed using KPCA than original sound recording feature space, where PD's new sound recording features has huge variance [31].

Consider $\{a_i\}, i = 1, \ldots N$ as Parkinson Disease (PD) dataset and D-dimensional sound recorded features vector is represented by every a_i. An M-dimensional sound reordered feature subspace is projected with PD data, where $< D$. Expression (1) represents projection as,

$$b = Xa,\ X = [u_1^T, \ldots u_M^T], u_k^T u_k = 1, k = 1, \ldots M \quad (1)$$

It requires maximization of $\{b_i\}$'s variance. So, X must be computed using expression (2) as,

$$X^* = \arg\max_X tr(S_b) \quad (2)$$

$$S_b = \frac{1}{N}\sum_{i=1}^{N}(b_i - \bar{b})\left(b_i - \bar{b}\right)^T \quad (3)$$

$$\bar{b} = \frac{1}{N}\sum_{i=1}^{N} a_i \quad (4)$$

Assume dataset $\{a_i\}$'s covariance matrix as S_a. So,

$$tr(S_b) = tr(XS_aX^T) \quad (5)$$

Lagrangian multipliers are used for computing expression (5) and final expression (6) is obtained by taking its derivative as,

$$S_a u_k = \lambda_k u_k \quad (6)$$

In shows that, S_a's eigenvector is u_k. Expression (7) is used for expressing a_i as,

$$a_i = \sum_{k=1}^{D}(a_i^T u_k)u_k \quad (7)$$

Expression (8) is used for approximating a_i as,

$$\tilde{a}_i = \sum_{k=1}^{M}(a_i^T u_k)u_k \quad (8)$$

Where, S_a's eigenvector is represented as u_k , which corresponds to kth largest eigen value. From original D-dimensional sound recorded features vector space, nonlinear transformation $\phi(a)$ is applied to forma Kernel matrix of M dimensional feature space from sound recorded feature vector, with $M < D$.

Then every Parkinson's Disease (PD) $(a_i)$is projected to a point $\phi(a_i)$. Consider zero mean of projected new sound recorded features using expression (9).

$$\frac{1}{N}\sum_{i=1}^{N} \phi(a_i) = 0 \quad (9)$$

The projected sound recorded features covariance matrix is M × M, which is computed as,

$$Co = \frac{1}{N}\sum_{i=1}^{N} \phi(a_i)\, \phi(a_i)^T \quad (10)$$

Expression (11) gives its eigenvectors an eigenvalues.

$$Cov_k = \lambda_k v_k \quad (11)$$

Where, k = 1, 2,….., M. From expression(10) and (11),

$$\frac{1}{N}\sum_{i=1}^{N} \phi(a_i)\, \{\phi(a_i)^T v_k\} = \lambda_k v_k \quad (12)$$

Using expression (13), it can be rewritten as,

$$v_k = \sum_{i=1}^{N} y_{ki}\phi(a_i) \quad (13)$$

Value of $v_k$is substituted in expression (12) with expression (13) and produced,

$$\frac{1}{N}\sum_{i=1}^{N} \phi(a_i)\phi(a_i)^T \sum_{j=1}^{N} y_{kj}\phi(a_i) = \lambda_k \sum_{i=1}^{N} y_{ki}\phi(a_i) \quad (14)$$

Sound recording's feature vector with reduced dimensionality is computed from this.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X
Received: 04 May 2020/Accepted: 05 June 2020

Then minimum Redundancy Maximum Relevance (mRMR) is used for selecting those features.

Feature Selection by Minimum Redundancy Maximum Relevance (mRMR)

For specified PD classification task, importance of PD features set is ranked using mRMR. According to relevance with target, PD features are ranked in this technique and also penalizes PD features redundancy.

Between class c and PD features set 'fe', computation of maximum dependency is a major objective and it uses Mutual Information (MI), which is represented as I. Expression (15) defines, MI between a PD features pair and it can be obtained effectively with known joint probability pr(x, y), marginal probabilities pr(x) and pr(y) between PD features pair.

$$I(X,Y) = \sum_{y\in Y}\sum_{x\in X} pr(x,y)\log\left(\frac{pr(x,y)}{pr(x)pr(y)}\right) \quad (15)$$

Maximum relevance has a search of those PD features, which satisfies following expression,

$$\max D(A,c)\,;\, D = \frac{1}{|A|}\sum_{A_i\in A} I(A_i,c) \quad (16)$$

Huge amount of redundancy may be produced, if PD features are selected based on maximum relevance criterion. So, expression (17) is added with minimum redundancy criterion as,

$$\min R(A)\,;\, R = \frac{1}{|A|}\sum_{A_i,A_j\in A} I\left(A_i,A_j\right) \quad (17)$$

The mRMR criterion is produced by combining and optimizing parameter R and D. Greedy algorithm is employed for the same in practice, where, selected PD features set is represented as S.

$$\max_{A_i\notin S}\left[I(A_i;c) - \frac{1}{|S|}\sum_{A_j\in S} I\left(A_j;A_j\right)\right] \quad (18)$$

Algorithm 1 presents procedure of original mRMR version. In this algorithm, between two elements, which may be an input PD features pair (line 7) or specified PD feature having class (line 4).

Between every PD features pair, computed this function despite the fact that, final classification is not having any affect of various PD features pair as they are irrelevant.

Algorithm 1: Feature Selection Algorithm based on Mrmr

Input: dataset$D$, selected features count $S_{fe}$from dataset 'D'
Output: finally selected features$St_{Fe}$from dataset 'D'
Start
For each feature Fe_iin Dataset (D)
do
Relevance =mutualInfo($Fe_i, class$)
Redundancy =0;
For each feature Fe_jin Dataset (D)
Do
Redundancy += mutualInfo (Fe_i, Fe_j);
End for
mRMR values[Fe_i] =relevance-redudancy;
End for
$St_{Fe} = sort$(mrmrvalues)$.found(S_{fe})$
End

PD Classification ViaFCLSTM-CNN Classifier

Fuzzy Convolution Long Short-Term Memory based Convolutional Neural Network (FCLSTM-CNN) is used for classifying PD in this study.

For PD classification, proposed FCLSTM-CNN classifier in this work. Introduced a FCLSTM-CNN model and figure 2 shows this classifier's architecture.

Fuzzy weights are computed in FCLSTM-CNN classifier using membership function and for extracting more relevant predicted Parkinson features, Parkinson features response are adjusted.

For step-by-step Parkinson classification, hidden states and Parkinson features are used in FCLSTM-CNN and in internal memory state, context data are maintained for mining correlation between PD features.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X
Received: 04 May 2020/Accepted: 05 June 2020

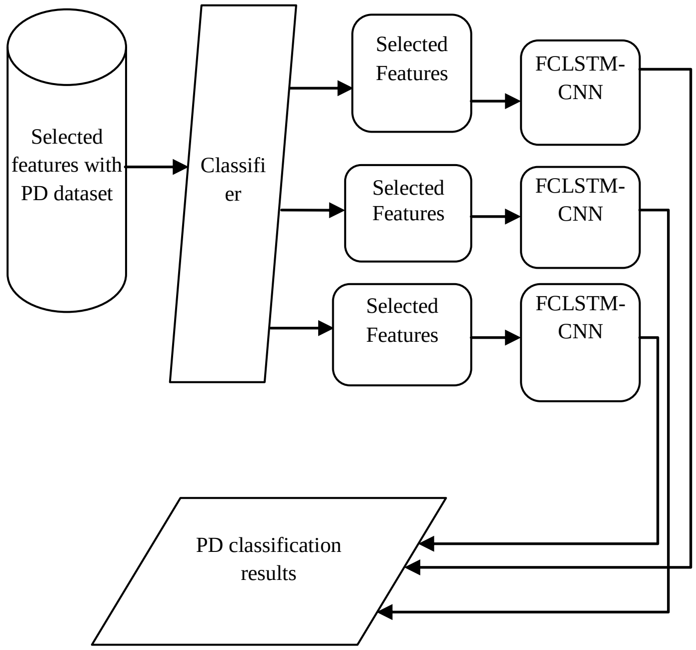


Figure 2: Architecture of Proposed Fuzzy Convolution Long Short-Term Memory Based Convolutional Neural Network (FCLSTM-CNN) Classifier

For exploring correlation between genes, FCLSTM is used in this work an genes are predicted one by one. In genes subsequent prediction, hidden state corresponds to CLSTM having historical data. In sequence modelling, better performance is shown by LSTM. In image processing, spatial data are ignored by LSTM in general. This is due to the fact that, in general LSTM classifier, data sequence is passed through a full connection layer and input data is flattened into a one-dimensional vector, which leads to data loss.

This is not conductive for Parkinson 's disease (PD) performance enhancement. In Parkinson's Disease (PD), for various attribute prediction, various levels of importance are shown by various image areas. For Parkinson's Disease (PD) performance enhancement, it is required to retain relevant spatial data.

In this classifier, CLSTM [32] is used for maintaining Parkinson's Disease (PD) genes structure. In FCLSTM, input is converted to state and state to state conversion are done using convolution operation. Attributes data can be captured in a better manner when compared with standard LSTM and between PD attributes, correlation can be captured in a better way. Following shows the expression of FCLSTM,

$$i_t = sigmoid(FW_{ix} * x_t + FW_{ih} * h_{t-1} + FW_{ic} \odot c_{t-1} + b_i) \quad (19)$$

$$f_t = sigmoid\left(FW_{fx} * x_t + FW_{fh} * h_{t-1} + FW_{fc} \odot c_{t-1} + b_f\right) \quad (20)$$

$$o_t = sigmoid(FW_{ox} * x_t + FW_{oh} * h_{t-1} + FW_{oc} \odot c_t + b_o) \quad (21)$$

$$g_t = tanh\left(FW_{gx} * x_t + FW_{gh} * h_{t-1} + b_g\right) \quad (22)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t, h_t = o_t \odot tanh(c_t) \quad (23)$$

Where, logistic sigmoid function is represented as *, hyperbolic tangent function is represented as $tanh(\cdot)$,tth step of CLSTM is represented using the subscript t, input gate is given by i_t, forget gate is given by f_t, output gate is given by o_t, input modulation gate is given by g_t, input data is given by x_t, cell state is given by c_t, hidden state is given by h_t. Three dimensional tensors are given by x_t, c_t, h_t, i_t, f_t and o_t. PD dataset is represented by first dimension and its rows and columns are represented by second and third dimensions. Every gates, fuzzy weight values are represented by $W_{ix}, FW_{ih}, FW_{ic}, FW_{fx}, FW_{fh}, FW_{fc}, FW_{ox}, FW_{oh}, FW_{oc}, FW_{gx}, FW_{gx}$. Fuzzy membership function having initial weights $W = (w_1, \dots w_i)$, i=1,...n is used for generating these weights, where, samples count of gene selected dataset is given by n.

Features PD dataset is maintained here using convolution operation. General LSTM and CLSTM core are essentially similar. Next layer input corresponds to previous layer output. With the inclusion of convolution operation, both spatial as well temporal relations can be acquired using CLSTM, which is a major difference with LSTM.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

This operation is same like spatial feature extraction on CNN convolution layer for obtaining features. In addition, data will effect convolutional operations because of stronger activating response of regions belonging to targeted pedestrian attribute.

As CLSTM is focusing on pedestrian attribute prediction's key areas, better performance is produced by it, when compared with general LSTM as shown in our experimentation. Figure 3, shows CLSTM's internal structure with Fuzzy weight values.

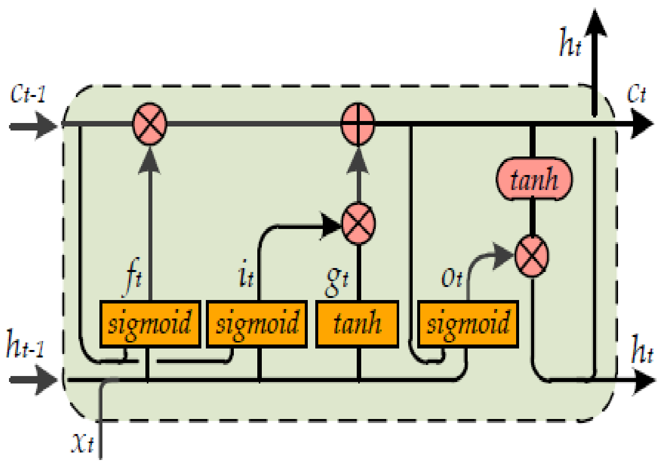


Figure 3:Internal Structure of Fclstm Classifier

Triangular membership function is used for computing fuzzy weights. An upper limit b, lower a and a value m, which lies between a and b are used for defining triangular function. Expression (24) defines the triangular membership function.

$$\mu_A(W) = \begin{cases} 0, W \leq a \\ \frac{W-a}{m-b}, a < W \leq m \\ \frac{b-W}{b-m}, m < W < b \\ 0, W \geq b \end{cases} \quad (24)$$

Generated fuzzy weights using this and above mentioned expressions from (19) to (23) are given with these values.

IV. Experimental Results

Details of obtained experimentation result in proposed FCLSTM-CNN classifier is explained in this section and methods like SCFW-KELM, SVM and CNN are used for making performance comparison. The LOPO CV is used for performing performance evaluation as instances in dataset are small. In every LOPO CV iteration, instances belonging to one individual is left out as a test set while remaining instances of others are used as a training set. Individuals class label is decided by considering class labels majority assigned to recordings as, every individual is having 3 recordings.

V. Evaluation Metrics

Classifier's predictability performances can be assessed using evaluation metrics. Most commonly used metric is accuracy and in data with unbalanced class distribution, it produces misleading results. Ability of classifier in distinguishing various classes, even with class imbalance is measured using a metrics like Matthews Correlation Coefficient and F-Measure.

Table 1 indicates the confusion matrix and for a binary classification, incorrectly and correctly classified instances count per class is indicated in that table. In that confusion matrix, counts of true negative (tn), false negative (fn), false positive (fp) and true positive (tp) are represented as tn, fn, fp and tp. F-Measure is computed as per these counts.

$$Precision = \frac{tp}{tp+fp} \quad (25)$$

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

$$recall = \frac{tp}{tp+fn} \quad (26)$$

$$F - measure = \frac{2*precision*recall}{precision+recall} \quad (27)$$

Binary classifications quality can be quantified using another metric called Matthews Correlation Coefficient (MCC). Counts of tn, fn,fp,tp are considered in MCC an it is assumed as a balanced measure as it can be used with unbalanced class distribution. Between predicted and actual instances, correlation coefficient is represented using MCC and its value lies between -1 to +1. Perfect prediction is indicated using +1 and disagreement between actual and predicted labels is represented with -1 value.

Table 1: Confusion Matrix For Two-Class Classification

ACTUAL/ PREDICTED AS	POSITIVE	NEGATIVE
POSITIVE	**Tp**	**Fn**
NEGATIVE	**Fp**	**tn**

VI. Results Comparison

Three features type and classifiers like existing SCFW-KELM, SVM, CNN and proposed FCLSTM-CNN classifiers are used in experimentation. Classifier performance is assessed using metrics like MCC< F-Measure and accuracy. For new experiments, employed triple feature sets in feature-level combination. Table 2 indicates the classification results.

Around 86.52**%** accuracy rate is produced with TQWT, MFCC and Wavelet features combination with 90.14**%** F-Measure rate, whereas, 83.6**%** accuracy rate is produced with TQWT+MFCC+Concat combination and 84.25**%** is produced with TQWT+Wavelet+Concat combination. With respect to MCC, F-Measure and accuracy scores, worst performance is exhibited by feature combinations without TQWT features (MFCC+Wavelet+Concat).

Table 2: Results of Feature-Level Combination: Triple Feature Sets

FEATURE COMBINATION	ACCURACY (%)	F-MEASURE (%)	MCC (%)
TQWT+MFCC+Wavelet	86.52	90.14	54.50
TQWT+MFCC+Concat	83.60	88.16	52.81
TQWT+ Wavelet + Concat	84.25	89.68	50.25
MFCC + Wavelet + Concat	82.48	87.41	40.28

Table 3: Results of SCFW-KELM Classifier with Triple Feature (KPCA+ mRMR)

FEATURE COMBINATION	ACCURACY (%)	F-MEASURE(%)	MCC(%)
TQWT+MFCC+Wavelet	71.9508	61.4472	50.4472
TQWT+MFCC+Concat	64.4472	70.9508	48.1725
TQWT+ Wavelet + Concat	69.9508	68.4472	48.4472
MFCC + Wavelet + Concat	75.2908	66.7772	52.6672

The SCFW-KELM classifier with MFCC + Wavelet + Concat features combination produces around 75.2908% of accuracy rate. In TQWT+ Wavelet + Concat, TQWT+ Wavelet + Concat and TQWT+MFCC+Wavelet performance will be enhanced as per models and features as mentioned in table 3. Triple feature set is used for training SVM. With respect to all metrics, among all classifiers, better performance is exhibited by MFCC + Wavelet + Concat combination in triple feature results as shown in figure 4. This combination produces 87.2293% accuracy rate, where 93.2752% of accuracy rate is produced by CNN classifier as mentioned in table 5. In MFCC+Wavelet+ Concat, TQWT+MFCC+Concat, TQWT+Wavelet+Concatand TQWT+MFCC+Wavelet, there will be an enhancement in performance according to CNN classifier models and features. Triple feature set is used for training SVM classifier. With respect to all metrics, among all classifiers, better performance is exhibited by MFCC + Wavelet + Concat combination in triple feature results as shown in table 4.

Table 4: Results of SVM Classifier with Triple Feature (KPCA+ mRMR)

FEATURE COMBINATION	ACCURACY (%)	F-MEASURE(%)	MCC(%)
TQWT+MFCC+Wavelet	84.2025	81.7152	54.9000
TQWT+MFCC+Concat	82.7640	79.5960	53.3000
TQWT+ Wavelet + Concat	86.4662	85.1589	57.5000
MFCC + Wavelet + Concat	87.2293	86.3510	58.4000

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

Table 5: Results of CNN Classifier With Triple Feature (KPCA+ mRMR)

FEATURE COMBINATION	ACCURACY (%)	F-MEASURE(%)	MCC(%)
TQWT+MFCC+Wavelet	86.6697	84.6697	55.9007
TQWT+MFCC+Concat	91.0315	89.0315	60.1656
TQWT+ Wavelet + Concat	87.1695	85.1695	62.0993
MFCC + Wavelet + Concat	93.2752	91.2752	63.2384

FCLSTM-CNN classifier in MFCC + Wavelet + Concatfeature combination produces 94.2557% of accuracy, 90.4729% of F-Measure rate and 65.9960% of MCC value, which is a greater one. Same accuracy values are produced by TQWT+Wavelet+Concat and TQWT+MFCC+Wavelet models as mentioned in table 6.

Table 6: Resultsof FCLSTM-CNN Classifier WITH Triple Feature (KPCA+ mRMR)

FEATURE COMBINATION	ACCURACY (%)	F-MEASURE(%)	MCC(%)
TQWT+MFCC+Wavelet	92.1457	93.0257	66.3669
TQWT+MFCC+Concat	92.1854	90.3252	66.4060
TQWT+ Wavelet + Concat	92.2252	92.2252	64.1457
MFCC + Wavelet + Concat	94.2557	90.4729	65.9960

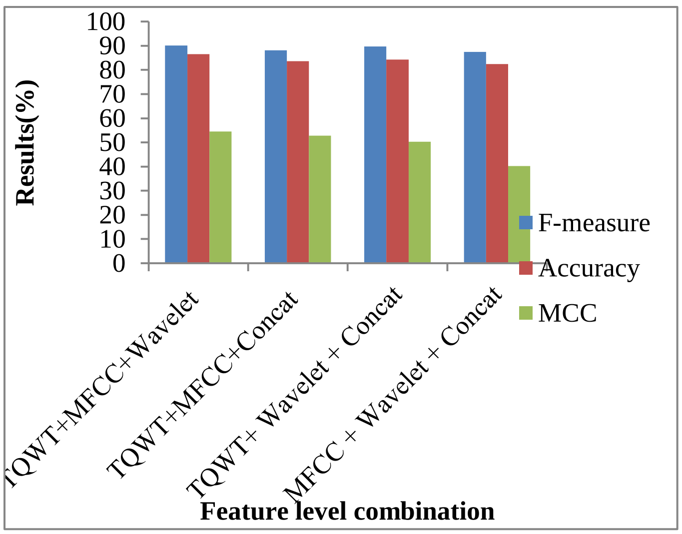


Figure 4:Feature Level Combination Results Vs. Metrics

Figure 4 indicates different metrics results of classifier with various feature set combination. Before the application classification task, these results are obtained. It indicates that TQWT+MFCC+Wavelet feature set produces 86.52% of accuracy, 90.14% of f-measure and 54.5%of MCC, when compared with other combination of feature sets.

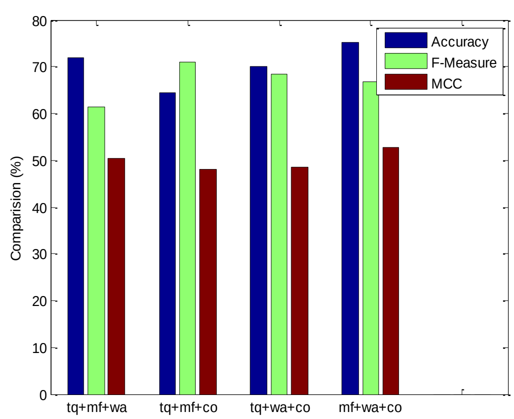


Figure 5: SCFW-Kelm Classifier with Feature Level Combination Results Vs. Metrics

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

Figure 5 indicates different metrics results of classifier based on SCFW-KELM classifier with various feature set combination.

It indicates that classifier model based on SCFW-KELM with MFCC + Wavelet + Concat feature set produces 75.2908% of accuracy, 66.7772% of f-measure and 52.6672% of MCC, when compared with other combination of feature sets.

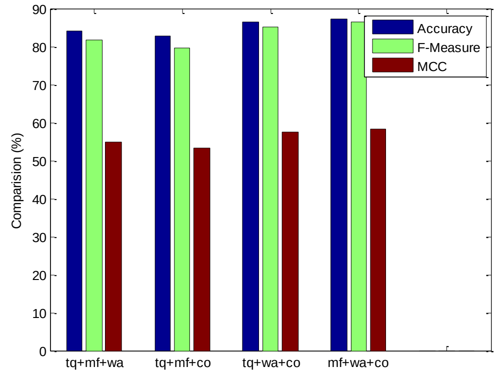


Figure 6: SVM Classifier WITH Feature Level Combination Results VS. Metrics

Figure 6 indicates different metrics results of classifier based on SVM classifier with various feature set combination. It indicates that classifier model based on SVM with MFCC + Wavelet + Concat feature set produces 87.2293%, of accuracy, 86.3510% of f-measure and 58.4000% of MCC, when compared with other combination of feature sets.

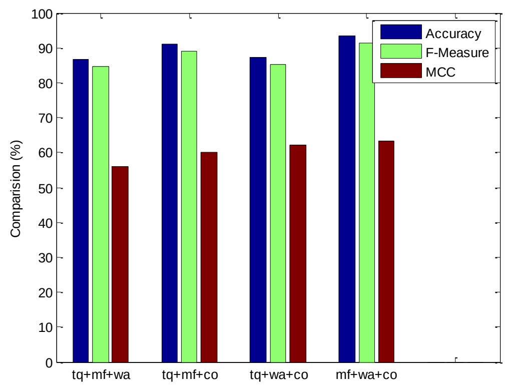


Figure 7: CNN Classifier WITH Feature Level Combination Results VS. Metrics

Figure 7 indicates different metrics results of classifier based on CNN model. It indicates that classifier model based on CNN with MFCC + Wavelet + Concat feature set produces 93.2752% of accuracy, 91.2752% of f-measure and 63.2384% of MCC, when compared with other feature sets.

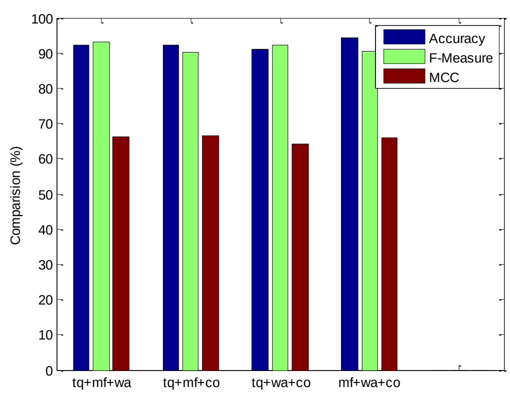


Figure 8: FCLSTM-CNN Classifier WITH Feature Level Combination Results VS. Metrics

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X
Received: 04 May 2020/Accepted: 05 June 2020

Figure 8 shows the results of various proposed FCLSTM-CNN classifier metrics. It indicates that proposed FCLSTM-CNN classifier with MFCC + Wavelet + Concat feature combination set produces 90.4729% of f-measure, 94.2557% of accurate and 65.9960% of MCC values, which are greater than other features sets. Optimum performance is produced by proposed work and using KPCA, dimensions of features are reduced.

VII. Conclusion and Future Work

Set of speech or vocal features are used for PD classification in this work and it proposed a deep FCLSTM-CNN classifier for the same. A CNN based framework is build for this purpose, which assist us in distinguishing PD patients and healthy individuals. For PD detection classification, extracted some features from input dataset. Kernel Principal Component Analysis (KPCA) is used for reducing the dimensionality of extracted features.

The minimum Redundancy Maximum Relevance (mRMR) is used for selecting informative features from entire features set. According to class label, features having high relevance score are selected by mRMR and redundant features are eliminated using this. According to the relevance between target and PD features, PD features are ranked by mRMR technique. It also penalizes PD features redundancy. At last, correlation between genes are explored using FCLSTM classifier and genes are predicted one by one. Dataset derived from University of California-Irvine (UCI) Machine Learning repository is used for training this classifier.

There are 252 individuals details in that dataset, which includes 64 healthy individuals and 188 PD patients. Three voice records per individual are included in that dataset. When compared to SCFW-KELM, SVM, CNN classifiers, highly accurate results are produced by proposed FCLSTM-CNN classifier as shown in results. Recent study can be extended in various ways in future. Various data types can be fed as an input network simultaneously using parallel convolution layers in proposed CNN. This creates an opportunity for utilizing multi-modal data in classification of PD. In classification process, various deep learning models can also be used in future.

References

[1] Jankovic, J., 2008. Parkinson's sdisease: clinical features and diagnosis. *Journal of neurology, neurosurgery* & psychiatry, 79(4), pp.368-376.

[2] Uebelacker, L.A., Epstein-Lubow, G., Lewis, T., Broughton, M.K. and Friedman, J.H., 2014. A survey of Parkinson's disease patients: most bothersome symptoms and coping preferences. *Journal of Parkinson's disease*, 4(4), pp.717-723.

[3] Skodda, S., 2011. Aspects of speech rate and regularity in Parkinson's disease. *Journal of the neurological sciences*, 310(1-2), pp.231-236.

[4] Bugalho, P. and Viana-Baptista, M., 2013. REM sleep behavior disorder and motor dysfunction in Parkinson's disease–a longitudinal study. *Parkinsonism & related disorders*, 19(12), pp.1084-1087.

[5] Reeve, A., Simcox, E. and Turnbull, D., 2014. Ageing and Parkinson's disease: why is advancing age the biggest risk factor?. *Ageing research reviews*, 14, pp.19-30.

[6] Sakar B. E., M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.

[7] Peker M., "A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and svm," *Journal of medical systems*, vol. 40, no. 5, p. 116, 2016.

[8] Sakar B. E., G. Serbes, and C. O. Sakar, "Analyzing the effectiveness of vocal features in early telediagnosis of parkinson's disease," *PloS one*, vol. 12, no. 8, p. e0182428, 2017.

[9] Lahmiri, S.; Shmuel, A. Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. *Biomed. Signal Process. Control* 2018, 49, 427–433.

[10] Gómez-García, J.A.; Moro-Velázquez, L.; Godino-Llorente, J.I. On the design of automatic voice condition analysis systems. Part I: Review of concepts and an insight to the state of the art. *Biomed. Signal Process. Control* 2019, 51, 181–199.

[11] Viteckova, S.; Kutilek, P.; Svoboda, Z.; Krupicka, R.; Kauler, J.; Szabo, Z. Gait symmetry measures: A review of current and prospective methods. *Biomed. Signal Process. Control* 2018, 42, 89–100.

[12] San-Segundo, R.; Navarro-Hellín, H.; Torres-Sánchez, R.; Hodgins, J.; De la Torre, F. Increasing Robustness in the Detection of Freezing of Gait in Parkinson's Disease. *Electronics* 2019, 8, 119.

[13] Kotsavasiloglou, C.; Kostikis, N.; Hristu-Varsakelis, D.; Arnaoutoglou, M. Machine learning-based classification of simple drawing movements in Parkinson's disease. Biomed. Signal Process. *Control* 2017, 31, 174–180.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

[14] Gallicchio, C.; Micheli, A.; Pedrelli, L. Deep Echo State Networks for Diagnosis of Parkinson's Disease. arXiv 2018, arXiv:1802.06708.

[15] Khatamino, P.; Cantürk, ˙I.;Özyılmaz, L. A Deep Learning-CNN Based System for Medical Diagnosis: An Application on Parkinson's Disease Handwriting Drawings. *In Proceedings of the 2018 6th International Conference Control Engineering Information Technology*, Istanbul, Turkey, 25–27 2018; pp. 1–6.

[16] Gil-Martín, M., Montero, J.M. and San-Segundo, R., 2019. Parkinson's disease detection from drawing movements using convolutional neural networks. *Electronics*, 8(8), p.907.

[17] Berus, L., Klancnik, S., Brezocnik, M. and Ficko, M., 2019. Classifying Parkinson's Disease Based on Acoustic Measures Using Artificial Neural Networks. *Sensors*, 19(1), pp.1-15.

[18] Haq, A.U., Li, J.P., Memon, M.H., Malik, A., Ahmad, T., Ali, A., Nazir, S., Ahad, I. and Shahid, M., 2019. Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. *IEEE Access*, 7, pp.37718-37734.

[19] Tsanas A., M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE Transactions on biomedical engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[20] KarimiRouzbahaniH. and M. R. Daliri, "Diagnosis of parkinson's disease in human using voice signals," *Basic and Clinical Neuroscience*, vol. 2, no. 3, pp. 12–20, 2011.

[21] Parisi L., N. RaviChandran, and M. L. Manaog, "Feature-driven machine learning to improve early diagnosis of parkinson's disease," *Expert Systems with Applications*, vol. 110, pp. 182–190, 2018.

[22] Wroge T. J., Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi, "Parkinson's disease diagnosis using machine learning and voice," in 2018 *IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE*, 2018, pp. 1–7.

[23] Gürüler H., "A novel diagnosis system for parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method," *Neural Computing and Applications*, vol. 28, no. 7, pp. 1657–1666, 2017.

[24] Ma, C.; Ouyang, J.; Chen, H.; Zhao, X. An Efficient Diagnosis System for Parkinson's Disease UsingKernel-Based Extreme Learning Machine with Subtractive Clustering Features Weighting Approach. *Comput. Math. Methods Med*. 2014, 2014, 985789.

[25] Pereira C. R., S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," *in 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE*, 2016, pp. 340–346.

[26] Eskofier B. M., S. I. Lee, J.-F. Daneault, F. N. Golabchi, G. Ferreira- Carvalho, G. Vergara-Diaz, S. Sapienza, G. Costante, J. Klucken, T. Kautz et al., "Recent machine learning advancements in sensor-based mobility analysis: Deep learning for parkinson's disease assessment*," In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE*, 2016, pp. 655–658.

[27] Gunduz, H., 2019. Deep Learning-Based Parkinson's Disease Classification Using Vocal Feature Sets. *IEEE Access*, 7, pp.115540-115551.

[28] Sakar C. O., G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, and H. Apaydin, "A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255–263, 2019.

[29] Yunusova Y., G. Weismer, J. R. Westbury, and M. J. Lindstrom, "Articulatory movements during vowels in speakers with dysarthria and healthy controls," *Journal of Speech, Language, and Hearing Research*, 2008.

[30] Selesnick I. W., "Wavelet transform with tunable q-factor," *IEEE transactions on signal processing*, vol. 59, no. 8, pp. 3560–3575, 2011.

[31] Wang, Q., 2012. Kernel principal component analysis and its applications in face recognition and active shape models. arXiv preprint arXiv:1207.3538.

[32] Xingjian, S.; Chen, Z.;Wang, H.; Yeung, D.;Wong,W.;Woo,W. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *In Proceedings of the Advances in Neural Information Processing Systems*, Montreal, QC, Canada, 2015; pp. 802–810.

DOI: 10.5373/JARDCS/V12SP7/20202108
ISSN 1943-023X

Received: 04 May 2020/Accepted: 05 June 2020

Hindawi
Journal of Healthcare Engineering
Volume 2022, Article ID 1487212, 12 pages
https://doi.org/10.1155/2022/1487212

Research Article

Optimization-Based Ensemble Feature Selection Algorithm and Deep Learning Classifier for Parkinson's Disease

B. Sabeena [1] **S. Sivakumari,** [1] **and Dawit Mamru Teressa** [2]

[1]*Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering, Coimbatore, India*
[2]*Department of Chemical Engineering, College of Biological and Chemical Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia*

Correspondence should be addressed to B. Sabeena; sabeena.bheeman@gmail.com and Dawit Mamru Teressa; dawit.mamiru@aastustudent.edu.et

Received 5 January 2022; Revised 4 February 2022; Accepted 22 February 2022; Published 13 April 2022

Academic Editor: Enas Abdulhay

PD (Parkinson's Disease) is a severe malady that is painful and incurable, affecting older human beings. Identifying PD early in a precise manner is critical for the lengthened survival of patients, where DMTs (data mining techniques) and MLTs (machine learning techniques) can be advantageous. Studies have examined DMTs for their accuracy using Parkinson's dataset and analyzing feature relevance. Recent studies have used FMBOAs for feature selections and relevance analyses, where the selection of features aims to find the optimal subset of features for classification tasks and combine the learning of FMBOAs. EFSs (ensemble feature selections) are viable solutions for combining the benefits of multiple algorithms while balancing their drawbacks. This work uses OBEFSs (optimization-based ensemble feature selections) to select appropriate features based on agreements. Ensembles have the ability to combine results from multiple feature selection approaches, including FMBOAs, LFCSAs (Lévy flight cuckoo search algorithms), and AFAs (adaptive firefly algorithms). These approaches select optimized feature subsets, resulting in three feature subsets, which are subsequently matched for correlations by ensembles. The optimum features are generated by OBEFSs the trained on FCBi-LSTMs (fuzzy convolution bi-directional long short-term memories) for classifications. This work's suggested model uses the UCI (University of California-Irvine) learning repository, and the methods are evaluated using LOPO-CVs (Leave-One-Person-Out-Cross Validations) in terms of accuracies, F-measure values, and MCCs (Matthews correlation coefficients).

1. Introduction

Parkinson's is a neurologic problem that involves tremors, rigidity, and problems moving, balancing, and coordinating. The signs of the disease normally appear slowly and continue to worsen. PD is a neurological malady classified as a motor system dysfunction. The patient's activities deteriorate with PD as it progresses. Patients are affected in their fundamental bodily systems, including breathing, balance, movements, and heart functioning [1], where, at initial stages, their speech flow gets hindered. The early diagnosis of PD leads to a longer life of patients, and the diagnostics require high precision and robust health informatics tools. Such solutions aim at assisting clinicians [2–4] who detect PD's severity using a range of sensors. This research work uses different speech signal processing methodologies to obtain PD's clinically relevant characteristics, which are then processed using learning algorithms to provide reliable detections of PDs.

The performances of computational algorithms are inextricably linked to the quality of input data. The manual identification of speeches or voices in a complex and intricate task can be executed efficiently by MLTs. Important features from voice signals can be identified by computer-based techniques, which may be one of the three categories, namely supervised, unsupervised, or semisupervised, based on the labeling of data. Filtering, wrapping, and embedding are the examples of supervised feature selection approaches.

Filtering strategies choose features that are unrelated to categorizations, while wrappers use the projected accuracies of previously determined values by algorithms for feature estimations. Embedded approaches like the filter models begin by selecting multiple potential feature subsets with specific cardinalities using statistical criteria, where subgroups with highest classification accuracies are finally selected. Unsupervised feature selections work on unlabeled data, however, evaluating the relevance of features is difficult for them. Using the labeled and unlabeled data and semi-supervised feature selections can evaluate feature relevance.

Computational methods based on biological evolutions provide a stronger basis for solving problems or taking decisions [5, 6]. EFSs boost the stability of feature selections as they take advantage of single approaches while overcoming their flaws. The analysis of features from datasets can be based on individual assessments or by the evaluation of subsets [7, 8]. Individual assessments create a rank of characteristics based on relevance, while alternative approaches employ search strategies to generate a series of feature subsets. These subsets are assessed iteratively using optimality criteria until they arrive at a final subset of selected characteristics [9]. This work's OBEFS framework guides the construction of EFSs that combine the benefits of several feature selection methods, avoid biases, and cover up their drawbacks.

The hierarchical layers of DNNs (deep neural networks), which are DLTs, manage to generate deep abstract representations of input features in applications. DLTs have been exploited in many applications, including speech recognition, image categorization, medication development, and genetic research [10]. Researchers have used DNNs for PD categorizations mainly because of their effectiveness [11, 12]. DNNs are very helpful classifiers in the case of PDs as they simulate complex and nonlinear data linkages. Previous research on PD classifications used single features like EEG data [11] and sensor activities [12] as inputs for CNNs (convolution neural networks), where the usage of unique parallel layers for classifications has not been tried. The study in [13] proliferated voices using more voice recordings of individuals in training and testing procedures with CVs (cross-validations), resulting in biased performance evaluations. Since the data had voice recordings of healthy persons and PD patients, LOPO-CVs were used to assess the performances of the proposed framework. LOPO-CVs removed examples from individuals in iterations in test sets while using other instances in training.

The suggested OBEFSs framework of this work selects features based on agreements. Instead of employing single feature selection approaches, the ensembles of feature selection methods aim to integrate numerous feature selection methods, such as FMBOAs, LFCSAs, and AFAs, whereas in OBEFSs, optimum features are utilized to train FCBi-LSTM classifiers. The proposed technique was trained using datasets from the UCI machine learning repositories, while its performance was validated using LOPO-CVs. This work's suggested model uses UCI learning repositories, and the methods are evaluated using LOPO-CVs in terms of accuracies, F-measure values, and MCCs.

2. Literature Review

In this part, we will outline some current works on PD classification that make use of machine learning techniques and discuss contemporary deep learning methods in PD classification. To evaluate speech recordings for PD classification, Alqahtani et al. [14] proposed classifications based on NNges (non-nested generalized exemplars), which, in spite of their capabilities, were not examined thoroughly. The study's experiments categorized healthy and PD using NNges and the algorithm's optimized parameters. Furthermore, the data was balanced using the synthetic minority oversampling technique (SMOTE) method. Finally, using the balanced data, NNge and ensemble algorithms, notably AdaBoostM1, were developed.

Using the sets of vocal data, Gunduz [15] used the dual frameworks of CNNs for identifying PDs, where different feature sets were generated but merged together. Their first architecture combined several feature sets before feeding them as inputs to 9-layered CNNs, while the second part fed feature set information directly to convolution layers in parallel. Hence, each parallel branch's deep features were obtained before their merger into layers. Their second showed highly promising results in tests as they learned deep features utilizing parallel convolutions. The extracted deep features were efficient in increasing the discriminative powers of classifiers in addition to differentiating patients with PDs from healthy people.

PDs were classified by Li et al. [16] by combining CART and ensemble learning. The study used CART to iteratively identify optimal training speech samples with high levels of differentiation. The study used ensembles, including RFs (random forests), SVMs (support vector machines), and ELMs (extreme learning machines) for learning optimal training data. The study classified test data using trained ensemble-learning systems. The study found that CART and RF combinations were stable when compared to other strategies and also improved PD predictions with speech data categorizations. Caliskan et al. [17] projected the diagnosis of PDs using speech impairments, the first indication of the disease. They used DNNs with stacked autoencoders and the softmax function for classifications. Their simulation results across two databases demonstrated the efficiency of DNN classifiers in comparison with other classification techniques.

For quickly detecting PDs, Cai et al. [18] proposed the usage of enhanced FKNNs (fuzzy K-nearest neighbors) combined with CBFOs (chaotic bacterial foraging optimizations) with Gauss mutations on voices data. Their CBFO-FKNN was an evolutionary instance-based learning methodology, where FKNN's parameters were tuned effectively by CBFOs. The study evaluated their suggested approach exhaustively on PD datasets in terms of classification accuracies, sensitivities, specificities, and AUCs (area under the receiver operating characteristic curves). The study aided physicians in making better clinical diagnostic judgments.

Castro et al. [19] classified PDs on UCI machine learning repository datasets with ANNs using MLPs (multilayer perceptrons). Their collections included voice recordings of patients with PDs along with control groups. The study used several networks and trained 10 to 6000 neurons, which were increased ten folds in the hidden layers. Their analyses of speech-related characteristics by ANNs could be used to assess patients' impacts of PDs. MLTs can identify other neurological disorders when biological data is made available. Disorders were classified by Abdurrahman and Sintawati, [20] where well-known speech characteristics were used in PD research, including jitters, shimmers, basic frequency parameters, and harmonicity parameters, and they assessed PDs using RPDEs (recurrence period density entropies), DFAs (detrended fluctuation analyses), and PPEs (pitch period entropies). PDs were classified using the XGBoost algorithm, which used identified baseline features, followed by feature selections executed from feature importance plots to enhance the model's performance. The resulting locShimmer features were eliminated from the model, and the efficacy of features was improved by XGBoost's assessments of feature importance to increase classification accuracies.

Karabayir et al. [21] examined PD data with multiple MLTs, including LGBs (light gradient boosts), EGBs (extreme gradient boosts), RFs, SVMs, KNNs, least absolute shrinkages, selection operator regressions, and LRs (logistic regressions). The study also conducted variable significance analyses to find important factors in people diagnosed with PDs. The study found that LGBs outperformed other MLTs in benchmarks and could be utilized to screen huge patient groups for PDs at low costs. Patra et al. [22] employed MLTs to assess the voices of patient datasets and identify PDs. The study's base classifiers were DTs (decision trees), LRs, and KNNs, which had their performances compared to ensembles like bagging, RFs, and boosts. Furthermore, the most important traits associated with classifications for PDs were discovered and prioritized, depending on feature importance with the aim of differentiating PD-affected patients by detecting dysphonia.

Parisi et al. [23] proposed the use of hybrid AIs (artificial intelligence) for examining the cases of PDs. The study used UCI's databases, where the dysphonic values of 68 patients' clinical ratings were considered for processing. The study's feature selections were based on MLP weights while ranking input features, where physiological and pathological patterns were given different weight values. This strategy reduced examinable features from 27 to 20, thus effectively reducing the dimensions for the learning of LSVMs (Lagrangian support vector machines). The proposed hybrid MLP-LSVMs performed well in benchmarks against the existing and previously proposed schemes and could be used in clinical environments for the detection of PDs.

Datasets with rich features were examined by Hasan and Hasan [24] using ANOVA (Analysis of Variance) F-score values to extract the top 50 features. Several MLTs were applied, and their results were compared to prior studies. Their experiments found that feeding select characteristics to RFs resulted in the greatest accuracy scores. Their use of ANOVA for feature extraction successfully retrieved important characteristics that distinguish PD patients from healthy persons while improving classification accuracy scores. Qasim et al. [25] suggested hybrid feature selection approaches for processing unbalanced PD datasets. SOMTE approach was used in the study to balance the dataset. Subsequently, RFEs (recursive feature eliminations) and PCAs (principal component analyses) were used to remove contradictions found in the dataset's features and reduce the processing times of PCAs. Their classifiers included bagging, KNNs, MLPs, and SVMs that worked on the acoustic recordings of PDs along with the patient's individual characteristics. Their idea of using SMOLTE with RFEs and PCAs in preprocessing datasets was also compared with other identifiers for PDs and general medical disorders found in people. The study was an asset to healthcare organizations.

Even though the first system integrates distinct selected features [15] prior to feeding them to a 9-layered CNN, the second model feeds feature sets to concurrent input layers that are directly connected to convolution layers. Before integrating deep features from each parallel connection in the merge layer, deep features from each parallel branch are extracted simultaneously. The suggested models are trained using information from UCI machine learning, and their results are verified using Leave-One-Person-Out Cross Validation (LOPO CV). The F-measure and Matthews correlation coefficient measure, as well as correctness, are employed to examine our data because of the imbalanced class distribution. This second model appears to be quite promising, as it is capable of learning feature representations from each set of features via concurrent convolution layers, according to experimental data.

3. Proposed Methodology

This research work proposes a new feature selection and classification framework for identifying PDs. This work uses five major steps, namely, the extraction of features based on voices, dimensionality reductions using KPCAs (kernel-based principal component analyses), the usage of proposed OBEFSs, LFCSAs, AFAs, and FCBi-LSTMs. Subsequently, the assessments are evaluated using LOPO-CVs. Figure 1 depicts the general flowchart of the proposed system.

3.1. PD Dataset. The PD dataset encompassed speech samples used by prior studies to diagnose PDs from UCI's machine learning repositories [13]. The data gathered at Istanbul University's Cerrahpasa Faculty of Medicine's Department of Neurology comprised 188 PD patients (107 men and 81 women) in the age range of [33, 87] and 64 healthy persons (23 men and 41 women) in [41, 82] age ranges. The voices were collected on 44.1 kHz (microphone's frequency), and three copies of the vowels of individuals were collected after doctor's examinations.

3.2. Feature Extractions. The dataset had baseline and temporal frequency features, MFCCs (Mel frequency cepstral coefficients), WTs (wavelet transforms), TQWTs (tunable Q-factor wavelet transforms), and vocal fold features:

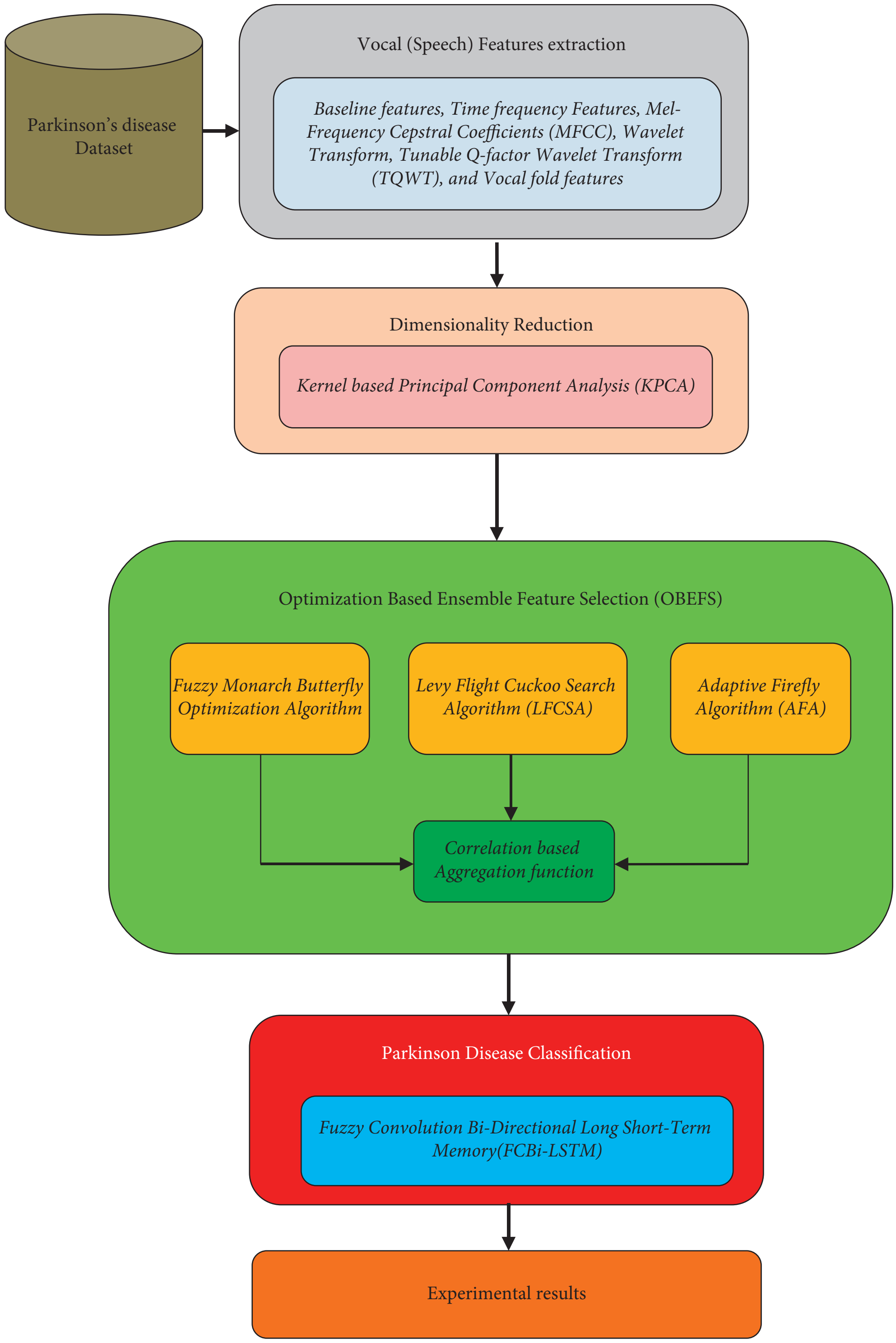


Figure 1: Overall flow of the proposed system.

(i) Baseline features: since PDs impede the speech of patients even in the early stages, speech characteristics were successfully used to evaluate PDs and track the disease's developments following medicinal therapies. The fundamental frequency parameters (#5), harmonicity parameters (#2), RTDEs (recurrence time density entropies) (#1), DFAs (detrended fluctuation analyses) (#1), and PPEs (#1) have been extensively utilized in characterizing speech-based PD researches [24, 26] and form the baseline features [13].

(ii) Time frequency features: intensity parameters (#3), formant frequencies (#4), and bandwidth (#4) are the examples of features.

(iii) MFCCs: MFCC-based extractions use triangular overlapped filter banks to combine cepstral analyses with spectral domain partitions. MFCCs can detect rapid deterioration in the movements of articulators in PDs like the tongues and lips, which are directly affected by the disease. The dataset had 84 characteristics related to MFCCs to identify the PD effects in the vocal tract (#84), and they were generated using the mean and standard deviation of initial 13 MFCCs along with the signal's log energies and $1^{st}/2^{nd}$ order derivatives [13], in addition to vocal folds.

(iv) WTs: generally**,** WTs are used to make decisions about signals and specifically on signals with minor fluctuations on regional scales. Several studies have utilized WT features obtained from a speech sample's raw fundamental frequencies (F0) to diagnose PD. This work produced 182 WTs characteristics from both approximations and detailed coefficients, including energies, Shannon's and log energy entropies, and Teager-Kaiser energies.

(v) TQWTs: the extraction of features using TQWTs improves signal qualities by adjusting three parameters, namely Q-factors (Q), redundancies (r), and a number of levels (J) based on the signal's behaviors. The oscillations in the time domain signals are proportional to Q-factors, while J stands for decomposed layer counts. On decompositions, J high-pass filters output $J + 1$ sub-bands and one final low-pass filtered output. Ringing, controlled by r allows wavelet's localizations with respect to time [27]. This study's tests yielded 432 TQWT-related characteristics from the dataset [13].

(vi) Vocal fold features: the effects of noises on vocal folds were also investigated in this work using features based on vocal fold vibrations. The study extracted the following from the data [13]: glottis quotients (GQs) (#3), glottal-to-noise excitations (GNEs) (#6), vocal fold excitation ratios (VFERs) (#7), and empirical mode decompositions (EMDs).

(vii) Concat features: concat features are the combination of baseline, vocal fold, and time frequency features.

3.3. Dimensionality Reduction Using KPCAs. Approaches based on KPCAs are prominent for dimensionality reductions. KPCAs consider linear subspaces with reduced dimensionalities in the original sound's feature spaces, where new sound recordings of PDs show the greatest variance in features [28]. Assuming $\{a_i\}, i = 1, \ldots N$ is the PD dataset, where a_i represents D-dimensional sound recorded feature vectors, they have to be projected into M-dimensional sound reordered feature subspaces that are lesser than D, and reduced feature vectors of sound recordings are identified. These reduced dimensional feature sets are used by OBEFSs for selecting relevant features.

3.4. Feature Selections Using OBEFSs. The proposed OBEFSs integrate the normalized results of multiple feature selections to arrive at quantitative feature sets with ensemble significances. In the initial phase, the series of feature selectors are created for different outputs, followed by the aggregations of a single model's results. The aggregations of feature selections are accomplished using correlations or consensus on feature ranks or counting most selected features for determining consensus-based feature subsets. The proposed OBEFSs generate final consensus ranks by combining feature ranks supplied by single feature selectors: FMBOAs, LFCSAs, and AFAs.

3.4.1. FMBOAs. This work uses FMBOAs for the selection of feature subsets, where the characteristics for samples are considered based on the effects of feature existences in PDs. Classifiers then use these selected attributes from samples (m denotes the number of voice samples). Classifiers forecast their own class labels, and evaluations are made for ultimate selections. The original characteristics are given feature weights that indicate their significance to classifications, and features with the highest weights are chosen. MBOs are migration-based that are built on migration trends, where fitness and importance of selections are rated. When used without modifications, FMBOAs show good classification accuracy results, indicating that they balance their global and local searches. The global search components of MBOAs were tweaked in this study to provide more precise results and boost effectiveness in locating the right characteristics before resorting to local searches. Individual butterflies analyze attributes that interact with one another on local levels, disseminating information across swarms and resulting in the system's growing capabilities [29–31]. They are carried out with the help of two operations, namely migration operators and adjustments to butterfly operators.

3.4.2. LFCSAs. CSAs (cuckoo search algorithms) are motivated by the unusual habit of cuckoo species, known as obligatory interspecific brood parasitism [32]. These behavioral patterns are based on the fact that certain animals use suitable hosts to optimize the selections of characteristics from datasets to grow their progenies. CSAs avoid parental commitments in rearing their offspring while limiting the dangers of egg loss (irrelevant traits) to other species. The final characteristics are chosen by placing eggs (features) in a variety of nests. The method's purpose is to replace the present solutions with eggs (irrelevant features) previously placed in the nest with these new solutions connected with cuckoo eggs (features). This iterative replacement may undoubtedly increase the quality of the solution over iterations, finally leading to a very good solution of the feature. In particular, CSA is based on three idealized rules [33, 34], which are as follows:

(1) Cuckoos lay the eggs (features) in nests randomly (accuracies).

(2) Nests with the best eggs (quality of features) are considered for subsequent generations for producing better solutions (features).

(3 The host nest counts are set with probability $\mathrm{prb}_a \in [0, 1]$. Hosts can find alien eggs (feature), a rule approximated by new nest replacements prb_a of the n available host nests. LFCSA algorithm initially begins with the N host. (1) gives the initial values of the k^{th} component of the j^{th} nest.

$$f_j^k(0) = \mu.\left(\mathrm{up}_j^k - \mathrm{low}_j^k\right) + \mathrm{low}_j^k, \tag{1}$$

where up_j^k is the k^{th} feature's upper bound, low_j^k is the k^{th} feature's lower bound, and μ is the uniform random variable in the range (0, 1). These parameters are adjusted for ensuring the feature values that exist with their feature spaces. The feature (egg), say i, randomly selected in the iteration, results in the solution f_i^{t+1}. The algorithm uses Lévy flights in place of random walks for efficient random searches. These flights, similar to random walks, are characterized by step sizes, following probability distributions with isotropic and random orientations. Lévy flights are depicted by

$$f_i^{t+1} = f_i^t + \alpha \oplus levy(\lambda). \tag{2}$$

The superscript t denotes the current generation, the symbol ⊕ denotes entry-wise multiplication, and $\alpha > 0$ denotes the step size. This step size specifies how far a particle (feature) may move in a certain number of iterations using a random walk. The Lévy distribution modulates the transition probability of the Lévy flights in

$$levy(\lambda) \sim g^{-\lambda}, \quad (1 < \lambda \le 3), \tag{3}$$

The production of random numbers with Lévy flights has two basic phases from a computational standpoint, which are as follows:

To begin, a random direction based on a uniform distribution is selected.

Then, based on the chosen Lévy distribution, a series of steps is constructed.

For symmetric distributions, Mantegna's approach is employed [34]. This method uses an equation to calculate the factor,

$$\widehat{\phi} = \left(\frac{\Gamma(1 + \widehat{\beta}).Sin(\pi.\widehat{\beta}/2)}{\Gamma\left((1 + \widehat{\beta}/2).\widehat{\beta}.2^{\widehat{\beta}-1/2}\right)} \right)^{1/\beta}, \tag{4}$$

where the Gamma function is denoted by Γ, and since $\widehat{\beta} = 3/2$ was utilized in a recent study [34], this work used the same ranges here. By (5), this factor is utilized in Mantegna's procedure to compute the step lengths:

$$\varsigma = \frac{u}{|v|^{1/\beta}}, \tag{5}$$

where u and v are the zero mean and deviation normal distributions σ_u^2 and σ_v^2, respectively. $\sigma_v = 1$ and σ_u follow the Lévy distribution given by (4). The step size ζ is then computed using

$$\varsigma = 0.01\varsigma\left(f - f_{best}\right). \tag{6}$$

The obtained ς changes the value of dimension x to: $f \leftarrow f + \zeta.\Psi$, where Ψ stands for the solution's random vector, and the x value lies in the normal distribution in the range (0, 1). LFCSA approaches identify new solutions (feature selections) that are fit (accurate) with existing solutions, where new solutions replace older ones on improvements. Nests with the worst values are discarded for further iterations and replaced with randomized new solutions, where replacement rates are based on probabilities prb_a, which are tuned for optimality. Thus, in iterations, existing solutions (feature selections) are rated based on their fitness values (accuracies), and the best solutions (features) are attained and stored as feature vectors f_{best}. Iterations are continued until the defined stopping criteria are met. LFCSA's pseudocode is depicted as Algorithm 1.

3.4.3. AFAs. The firefly algorithm is based on the idealized behavior of firefly flashing [35]. For the core formulation of FA, the three rules idealized are as follows:

(i) Because all fireflies are unisex, they will attract each other regardless of their gender for the best feature selection from the dataset

(ii) The brightness (accuracy) of a firefly is related to its attractiveness, which decreases as the distance between two fireflies grows

(iii) The brightness of a firefly is controlled by the objective function (accuracy)

The light intensity (In) varies exponentially and monotonically with distance. Equation (7) is used to explain it.

$$\mathrm{In} = \mathrm{In}_0 e^{-\gamma r}, \tag{7}$$

where In_0 is the initial light intensity and γ is the light absorption coefficient. As a firefly's attractiveness is proportional to the light intensity seen by neighbor fireflies (features), define the attractiveness β of a firefly by

$$\beta = \beta_0 e^{-\gamma r^2}, \tag{8}$$

where $\beta_0 = 1$ is the attractiveness at $r = 0$. The movement of a firefly (feature) "i" is attracted to another more attractive firefly(feature) "j", which is determined by

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}\left(x_i - x_j\right) + \alpha\varepsilon. \tag{9}$$

The third term is the randomization with the step α, being drawn from a Gaussian distribution.

FAs generically use (9) for iterative randomizations, resulting in uniform distributions in the interval [0, 1] range. Their step determinations are static/linear and are defined for unchangeable maximum generations. FAs begin with the same steps, and their values keep decreasing in iterations. As a result, it is possible that it will get stuck at the local optimum, causing premature convergence. Secondly, taking such a large stride may lead the firefly to miss the best option while it is still in the area of the firefly during the early phases of the search. As a result, search performance might be harmed.

Thus, (9) implies the benefits of explorations in FAs, where larger steps result in global optimum convergences.

(1) Begin
(2) Objective function $f(f)$, $f = (f_1, \ldots f_d)^T$ with $d = \dim(\Omega)$
(3) Generate initial population of N host nests $f_i \, (i = 1, \ldots N)$
(4) While ($t <$ MaxGeneration)
(5) Get a cuckoo (say i) randomly by Lévy flights
(6) Evaluate fitness F_i by the accuracy of the classifier
(7) Choose a nest among N (say j) randomly
(8) **If** $(F_i > F_j)$
(9) Replace j with the new solution
(10) End
(11) A fraction (prb_a) of worse nest(features in the dataset) are abandoned and new features are built by
(12) Lévy Flights
(13) Keep the best solutions using the accuracy of the feature
(14) End while
(15) Postprocess the results and visualization
(16) End

Algorithm 1: Levy flights cuckoo search algorithm.

For steps with low values, considerable influence occurs on explorations and convergences of algorithms. The values keep declining slowly on more iterations, however, they are faster in reduced iterations. These issues have been overcome in this study by the usage of self-adaptive steps, where the firefly's unique experiences help in selecting the best features from the data.

Step settings should be used to remedy the difficulties listed above. The firefly step should be set to be far away from the ideal solution. Fireflies between the two are utilized to balance the global and local searches for the best feature selection from the dataset. As a result, the firefly's stride must be concerned with both its previous data and current circumstances. This work introduces the firefly's history data, which contains the optimal value of the previous two iterations. Based on the comments mentioned above and many experiments, the step α of each firefly is calculated by (10) and (11), respectively. It is discussed as follows:

$$h_i(t) = \frac{1}{\sqrt{\left(f_{\mathrm{pi}}(t-1) - f_{\mathrm{pi}}(t-2)\right)^2 + 1}}, \quad (10)$$

$$\alpha_i(t+1) = 1 - \frac{1}{\sqrt{\left(f_{\mathrm{best}}(t) - f_i(t)\right)^2 + h_i(t)^2 + 1}}, \quad (11)$$

where $h_i(t)$ is the past two iterations' history data of the i^{th} firefly. f_{pi} is the fitness value of the best solution of the i^{th} firefly. f_{best} is the fitness value of the best solution of population heretofore found, and f_i is the fitness value of the i^{th} firefly, which reflects the current data. The firefly's next iterations are self-adaptive and are decided by the gap between the current fitness values and the population's best fitness values. As a result, the firefly steps might change with repetitions, and each firefly's step is, likewise, changed at the same time.

(1) Begin
(2) Objective function $f(x)$, $x = (x_1, \ldots x_d)^T$
(3) Generate initial population of n fireflies $x_i \, (i = 1, \ldots n)$
(4) Formulate light intensity In by objective function $f(x)$
(5) While ($t <$ MaxGeneration)
(6) Define absorption coefficient γ
(7) Evaluate fitness F_i by accuracy of the classifier
(8) For $i = 1$ to n(n fireflies)
(9) For $j = 1$ to n(n fireflies)
(10) If $(I_j > I_i)$
(11) Move firefly i towards j
(12) End if
(13) Vary attractiveness with distance r via exp $(-\gamma r^2)$
(14) Evaluate new selected features solutions and update light intensity
(15) Update the step of each firefly. The step is calculated by (10) and (11).
(16) End for j
(17) End for i
(18) Rank the best features and find the current best features
(19) End *while*
(20) Postprocessing the results and visualization
(21) End

3.4.4. Correlation Function. Correlations between the features are computed by ensemble feature selectors, where high similarities between the features award their eliminations. The features selected using three procedures form ensemble features, where only ideal feature sets are selected by majority votes and based on the outputs of individual feature sets. The correlation coefficient matrices for the features selected in the out-ensemble feature selection outputs are computed using

$$\text{correlation coefficient} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}}, \tag{12}$$

where x and y are the attribute values under consideration, and N is the total number of instances. The feature set selected by the correlation-based ensemble feature selector is given as an input to the classification.

3.5. Classification of PDs Using FCBi-LSTMs. This work used FCBi-LSTMs for the classification of PDs. The suggested approach computes fuzzy weight with membership values that are adjusted for extracting the most relevant features with respect to PDs. FCBi-LSTMs and CNNs analyze the selected characteristics from PD datasets [36]. CNNs made of convolution and pooling layers convolute and pool where outputs are fed to subsequent convolution layers. CNNs offer significant advantages in terms of feature extractions as they use partial filters for convolutions based on their understanding of biological vision cells' local perception. The convolution layer is separated into many output matrices using filters to offer a better representation of the selected features from the PD dataset, with each output matrix having a size of $(Nm + 1)$. The pooling layer of CNN is a technique for reducing the dimension of a matrix while keeping the fundamental links between the features. Pooling layers are average pooling layers with inputs from convolution layers. In the Bi-LSTM data analysis technique, the output of the last convolution layer is used as an intermediate variable [37]. As a result, LSTM does more than just add a nonlinear element to the input and loop cell transformation. Fuzzy weights are computed using Gaussian membership functions, where Bi-LSTMs outperform unidirectional LSTMs as they capture more structural information. The final outputs of Bi-LSTMs are processed by CNN's convolution layers for diagnosing PDs. To combine features processed by CNN and features processed by Bi-LSTM, multimodal factorized bilinear pooling (MFB) is utilized.

4. Experimental Results

This section describes the experimental findings achieved by the proposed FCBi-LSTM classifier and compares them to approaches, such as FCLSTM-CNN (fuzzy convolution long short-term memory-based convolution neural networks), CNN, and SVM. Since the samples in the test sets were fewer, LOPO-CVs' performance was evaluated using the training set's remaining individual instances, as each individual had three recordings, and the class labels assigned to these recordings were used to establish the individual's class label. The MIT-BIH arrhythmia database was used to conduct the investigations on arrhythmia recognition and classification systems and MATrix LABoratory R2016a (MATLAB R2016a). The implementation has been done using the following system specifications: Intel (R) Core™i3-4160T CPU@3.10 GHz 3.09 GHz processor, 4.00 GB RAM, Windows 8.1 Pro, 64-bit operating system, and 1 TB hard disk.

4.1. Evaluation Metrics. To test the predictability of the classifiers, evaluation metrics are required. Although accuracy is a widely used statistic, it might produce deceptive findings when data has an imbalanced class distribution. Even when there is a class imbalance, evaluation measures like F-measure and MCCs may be used to assess how effectively a classifier can discriminate between distinct classes. Allow the confusion matrix in Table 1 to represent the numbers of properly and erroneously categorized occurrences per class for binary classification. The letters tp, fp, fn, and tn in the confusion matrix mean true positive (tp), false positive (fp), false negative (fn), and true negative (tn), respectively. Precision, recall, F-measure, accuracy, and error were calculated using the formulae based on these counts.

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}, \tag{13}$$

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}, \tag{14}$$

$$F - \text{measure} = \frac{2^{*}\text{precision}^{*}\text{recall}}{\text{precision} + \text{recall}}, \tag{15}$$

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}}, \tag{16}$$

$$\text{error} = 100 - \text{Accuracy}. \tag{17}$$

MCCs, which take into consideration the tp, fp, fn, and tn counts and are frequently recognized as a balanced measure that may be employed even if the class distribution is uneven, are another statistic for evaluating the validity of binary classifications. MCCs are simply correlation coefficients ranging from −1 to +1 between the actual and predicted occurrences. A score of +1 indicates a perfect prediction, whereas a value of −1 indicates a discrepancy between the forecast and the actual labeling.

4.2. Results Comparison. Experimental evaluations of classifiers were executed with three types of features in terms of accuracy, error, F-measure, and MCC. The combination of MFCCs + Wavelets + Concated features with SVM resulted in the accuracy rate of 88.1294%, although the accuracy rate of MFCCs + Wavelets + Concated combination was 94.1752% for CNN. FCLSTM-CNN had the accuracy results of 93.0470%, 93.0854%, 93.1261%, and 95.1557%, respectively, for TQWT + MFCC + Wavelet, TQWT + Wavelet + Concat, TQWT + MFCC + Concat, and MFCC + Wavelet + Concat. The suggested FCBi-WLSTM classifier with MFCCs + Wavelets + Concated combinations achieved the highest accuracy rates of 98.7720% (F-measure rate of 98.5010% and 71.400% for MCC) (See Table 2).

Figures 2–5 show the F-measures, accuracies, MCCs, and errors of feature set combinations, where the TQWT + MFCC + Wavelet combination of feature sets yielded higher results of 98.3100 percent, 96.6381 percent, 74.300 percent, and 3.3619 percent for f-measures,

Table 1: Confusion matrix for two-class classification.

Actual/predicted as	Positive	Negative
Positive	tp	fn
Negative	fp	tn

Table 2: Results of classifiers with triple feature (KPCA + OBEFS).

Feature combination	F-measure	Accuracy	Error	MCC
SVM classifier (%)				
TQWT + MFCC + Wavelet	82.9150	85.1035	14.8965	56.2000
TQWT + MFCC + Concat	80.7960	83.6640	16.3360	54.6000
TQWT + Wavelet + Concat	86.3590	87.4662	12.5338	58.8000
MFCC + Wavelet + Concat	87.4510	88.1294	11.8706	59.7000
CNN classifier (%)				
TQWT + MFCC + Wavelet	85.8697	87.5696	12.4304	57.2007
TQWT + MFCC + Concat	90.2315	91.9315	8.0684	61.4600
TQWT + Wavelet + Concat	86.3695	88.0694	11.9306	63.3994
MFCC + Wavelet + Concat	92.4752	94.1752	5.8248	64.5384
FCLSTM-CNN classifier (%)				
TQWT + MFCC + Wavelet	94.2258	93.0470	6.9530	67.6669
TQWT + MFCC + Concat	91.5250	93.0854	6.9146	67.7060
TQWT + Wavelet + Concat	93.4200	93.1261	6.8739	65.4457
MFCC + Wavelet + Concat	91.6921	95.1557	4.8443	67.2960
FCBi-LSTM classifier (%)				
TQWT + MFCC + Wavelet	98.3100	96.6381	3.3619	74.300
TQWT + MFCC + Concat	96.5900	98.0244	1.9756	72.300
TQWT + Wavelet + Concat	97.5200	97.3457	2.6543	70.300
MFCC + Wavelet + Concat	98.5010	98.7720	1.2280	71.400

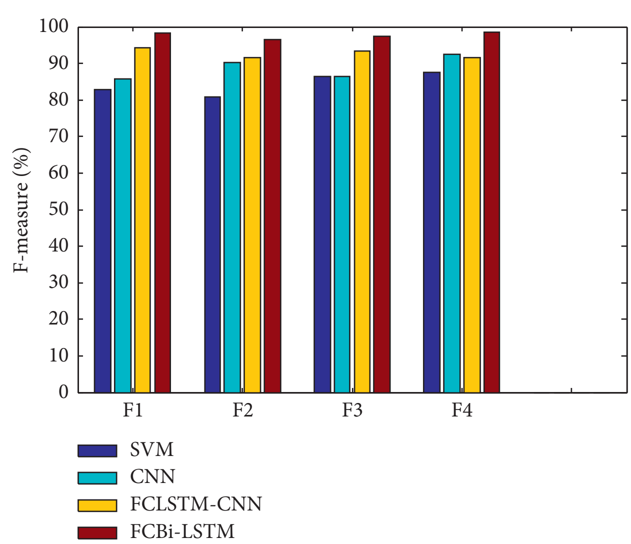


Figure 2: F-measure results of feature level combination vs. classifiers.

accuracies, MCCs, and errors, respectively, when compared to other combinations. Figure 2 compares the F-measure outcomes of four distinct feature level combinations using various classifiers. The proposed FCBi-LSTM with the first feature level combination achieved a higher F-measure value of 98.3100%, which was better than SVM, CNN, and FCLSTM-CNN, which achieved F-measures of 82.9150 percent, 85.8697 percent, and 94.2258 percent, respectively, at the first feature level combination.

Figure 3 depicts accuracies in the x-axis assessed using feature-level combinations on classifiers. FCBi-LSTM achieved 98.772 percent accuracy when compared to SVM, CNN, and FCLSTM-CNN, which achieved 88.1294 percent, 94.1752 percent, and 95.1557 percent accuracy, respectively, at the final feature level combination.

Figure 4 depicts error result comparisons of classifiers with four distinct feature level combinations. According to Figure 4, FCBi-LSTM results on final feature level

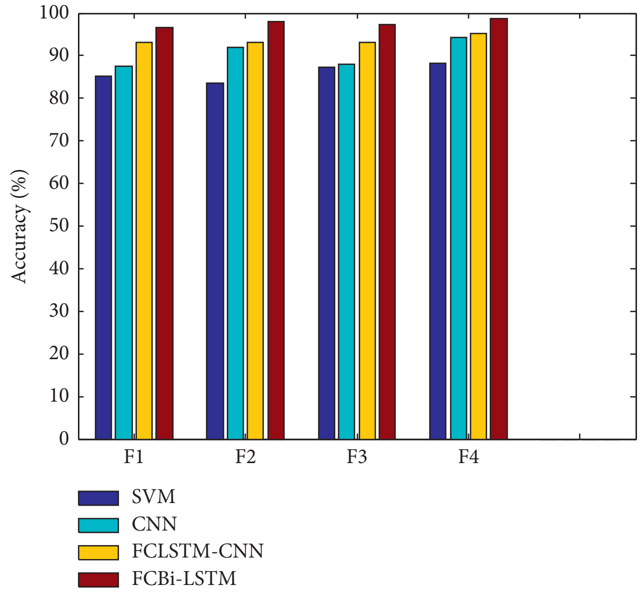


Figure 3: Accuracy results of feature level combination vs. classifiers.

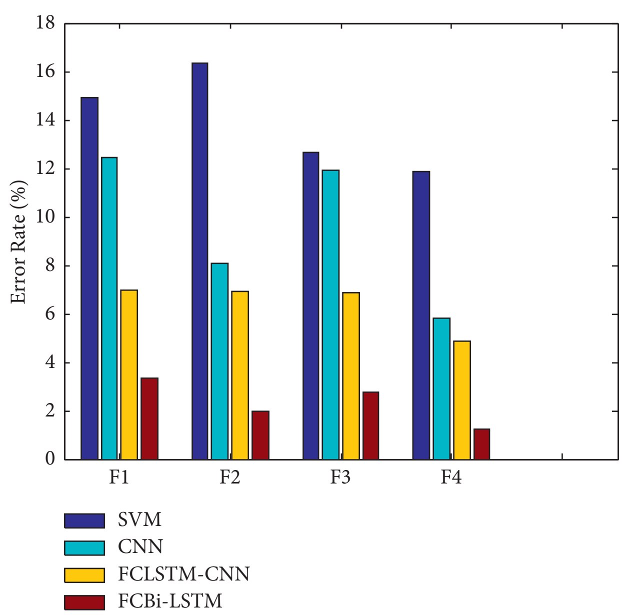


Figure 4: Error results of feature level combination vs. classifiers.

combinations produced reduced error values of 1.2280 percent, whereas SVM, CNN, and FCLSTM-CNN had higher error values of 11.8706 percent, 5.8248 percent, and 4.8443 percent, respectively, at the final feature level combination.

Figure 5 compares MCC results of TQWT + MFCC + Concat feature set, which yields a higher result of 74.30 percent for the proposed method, and 56.20 percent, 57.2007 percent, and 67.6669 percent for SVM, CNN, and FCLSTM-CNN classifiers, respectively (1st feature level combination). Because feature selection is accomplished using the proposed approach achieves superior MCC outcomes for all classifiers (OBEFSs).

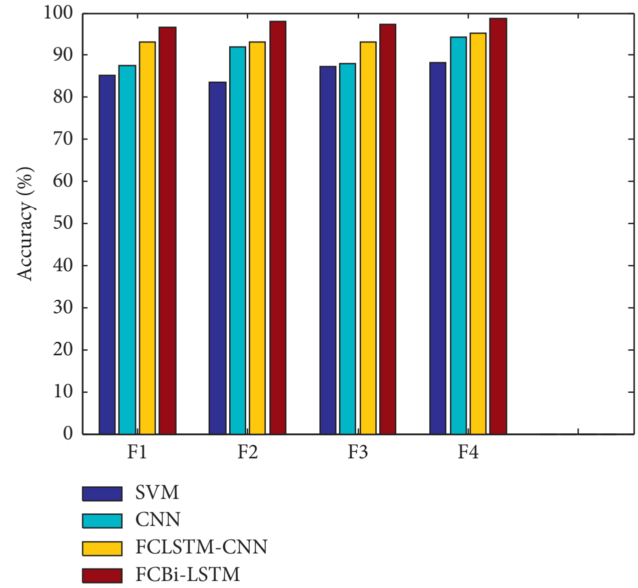


Figure 5: MCC results of feature level combination vs. classifiers.

5. Conclusion and Future Work

PD is the second most prevalent neurological ailment, causing considerable impairment, lowering the quality of life, and having no treatment. It is critical to diagnose PD early to use neuroprotective and early treatment techniques. In this research, a feature selection is used to present a multiclass classification challenge for PD analysis. For PD analysis, OBEFS and FCBi-LSTM are presented. The proposed OBEFS method is based on a number of algorithms, including FMBOA, LFCSA, and AFA. To execute OBEFS, the correlation function is utilized to choose optimum features from the three feature subsets. The FCBi-LSTM classifier is then used for PD diagnosis. It is an effective and accurate model for properly diagnosing the condition at an early stage, which might help doctors aid in the cure and recovery of PD patients. Classification algorithms were tested with UCI's machine learning libraries, and their performance is measured using precision, recall, F-measure, accuracy, and MCC. The results were compared to other existing techniques, and the findings show that the suggested model's accuracy is higher than the other current approaches. Deep learning has a bright future in engineering and medicine. In terms of future work, the goal is to extend existing research in novel ways. Different data types can be sent into the network as inputs at the same time using the proposed CNN's parallel convolution layers. It gives us the chance to utilize the multimodal data in PD classification. Also, the authors plan to use different deep learning models in the classification process.

Abbreviations

PD: Parkinson's Disease
DMTs: Data mining techniques
MLTs: Machine learning techniques
EFSs: Ensemble feature selections

OBEFSs: Optimization-based ensemble feature selections
LFCSAs: Lévy Flight Cuckoo Search Algorithms
AFAs: Adaptive firefly algorithms
FCBi-LSTMs: Fuzzy convolution bidirectional long short-term memories
UCI: University of California-Irvine
LOPO-CVs: Leave-One-Person-Out-Cross Validations
MCCs: Matthews correlation coefficients.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] T. V. S. Sriram, M. V. Rao, G. V. S. Narayana, and D. S. V. G. K. Kaladhar, "Diagnosis of Parkinson disease using machine learning and data mining systems from voice dataset," in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications*, pp. 151–157, (FICTA), Berlin, Germany, January 2014.
[2] H. Gürüler, "A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method," *Neural Computing & Applications*, vol. 28, no. 7, pp. 1657–1666, 2017.
[3] M. Peker, "A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM," *Journal of Medical Systems*, vol. 40, no. 5, pp. 116–16, 2016.
[4] B. E. Sakar, G. Serbes, and C. O. Sakar, "Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease," *PloS one*, vol. 12, no. 8, pp. 1–18, 2017.
[5] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, 2015.
[6] F. Moslehi and A. Haeri, "An evolutionary computation-based approach for feature selection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 9, pp. 3757–3769, 2020.
[7] B. Seijo-Pardo, V. Bolón-Canedo, I. Porto-Díaz, and A. Alonso-Betanzos, "Ensemble feature selection for rankings of features," *Advances in Computational Intelligence*, in *Proceedings of the International Work-Conference on Artificial Neural Networks*, pp. 29–42, Springer, Palma de Mallorca, Spain, June 2015.
[8] D. Guan, W. Yuan, Y.-K. Lee, K. Najeebullah, and M. K. Rasel, "A review of ensemble learning based feature selection," *IETE Technical Review*, vol. 31, no. 3, pp. 190–198, 2014.
[9] M. Mera-Gaona, D. M. López, R. Vargas-Canas, and U. Neumann, "Framework for the ensemble of feature selection methods," *Applied Sciences*, vol. 11, no. 17, pp. 1–16, 2021.
[10] S. L. Oh, Y. Hagiwara, U. Raghavendra et al., "A deep learning approach for Parkinson's disease diagnosis from EEG signals," *Neural Computing & Applications*, vol. 32, no. 4, pp. 1–7, 2020.
[11] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics," in *Proceedings of the 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 340–346, (SIBGRAPI), Sao Paulo, Brazil, October 2016.
[12] B. M. Eskofier, S. I. Lee, J.-F. Daneault et al., "Recent machine learning advancements in sensor-based mobility analysis: deep learning for Parkinson's disease assessment," in *Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 655–658, (EMBC), Orlando, FL, USA, Augeust 2016.
[13] C. O. Sakar, G. Serbes, A. Gunduz et al., "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255–263, 2019.
[14] E. J. Alqahtani, F. H. Alshamrani, H. F. Syed, and S. O. Olatunji, "Classification of Parkinson's disease using NNge classification algorithm," in *Proceedings of the 2018 21st Saudi Computer Society National Computer Conference (NCC)*, Riyadh, Saudi Arabia, April 2018.
[15] H. Gunduz, "Deep learning-based Parkinson's disease classification using vocal feature sets," *IEEE Access*, vol. 7, pp. 115540–115551, 2019.
[16] Y. Li, L. Yang, P. Wang et al., "Classification of Parkinson's disease by decision tree based instance selection and ensemble learning algorithms," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 2, pp. 444–452, 2017.
[17] A. Caliskan, H. Badem, A. Basturk, and M. E. Yuksel, "Diagnosis of the Parkinson disease by using deep neural network classifier," *IU-Journal of Electrical & Electronics Engineering*, vol. 17, no. 2, pp. 3311–3318, 2017.
[18] Z. Cai, J. Gu, C. Wen et al., "An intelligent Parkinson's disease diagnostic system based on a chaotic bacterial foraging optimization enhanced fuzzy KNN approach," *Computational and mathematical methods in medicine*, vol. 2018, pp. 1–24, Article ID 2396952, 2018.
[19] C. Castro, E. Vargas-Viveros, A. Sánchez, E. Gutiérrez-López, and D.-L. Flores, "Parkinson's disease classification using artificial neural networks," *IFMBE Proceedings*, in *Proceedings of the Latin American Conference on Biomedical Engineering*, pp. 1060–1065, Springer, Cancun, Maxico, October 2019.
[20] G. Abdurrahman and M. Sintawati, "Implementation of xgboost for classification of Parkinson's disease," in *Journal of Physics: Conference Series*, vol. 1538, no. 1, p. 012024, IOP Publishing, 2020.
[21] I. Karabayir, S. M. Goldman, S. Pappu, and O. Akbilgic, "Gradient boosting for Parkinson's disease diagnosis from voice recordings," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–7, 2020.
[22] A. K. Patra, R. Ray, A. A. Abdullah, and S. R. Dash, "Prediction of Parkinson's disease using ensemble machine learning classification from acoustic analysis," in *Journal of Physics: Conference Series*, vol. 1372, no. 1, IOP Publishing, Article ID 012041, 2019.
[23] L. Parisi, N. RaviChandran, and M. L. Manaog, "Feature-driven machine learning to improve early diagnosis of Parkinson's disease," *Expert Systems with Applications*, vol. 110, pp. 182–190, 2018.
[24] K. A. Hasan and M. A. M. Hasan, "Classification of Parkinson's disease by analyzing multiple vocal features sets," in

Proceedings of the 2020 IEEE Region 10 Symposium (TEN-SYMP), pp. 758–761, Dhaka, Bangladesh, June 2020.

[25] H. M. Qasim, O. Ata, M. A. Ansari, M. N. Alomary, S. Alghamdi, and M. Almehmadi, "Hybrid feature selection framework for the Parkinson imbalanced dataset prediction problem," *Medicina*, vol. 57, no. 11, pp. 1–13, 2021.

[26] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[27] I. W. Selesnick, "Wavelet transform with tunable q-factor," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3560–3575, 2011.

[28] Q. Wang, "Kernel principal component analysis and its applications in face recognition and active shape models," 2012.

[29] M. Alweshah, "Solving feature selection problems by combining mutation and crossover operations with the monarch butterfly optimization algorithm," *Applied Intelligence*, vol. 51, no. 1, 2020.

[30] G GWang, X. Zhao, and S. Deb, "A novel monarch butterfly optimization with greedy strategy and self-adaptive," in *Proceedings of the Soft Computing and Machine Intelligence (ISCMI), 2015 Second International Conference on*, pp. 45–50, Hong Kong, China, November 2015.

[31] Y. Feng, G.-G. Wang, W. Li, and N. Li, "Multi-strategy monarch butterfly optimization algorithm for discounted {0-1} knapsack problem," *Neural Computing & Applications*, vol. 30, no. 10, pp. 3019–3036, 2018.

[32] A. Mesa, K. Castromayor, C. Garillos-Manliguez, and V. Calag, "Cuckoo search via Levy flights applied to uncapacitated facility location problem," *Journal of Industrial Engineering International*, vol. 14, no. 3, pp. 585–592, 2018.

[33] M. Guerrero, O. Castillo, and M. García, "Cuckoo search via Lévy flights and a comparison with genetic algorithms," in *Fuzzy Logic Augmentation of Nature-Inspired Optimization Metaheuristics*, pp. 91–103, Springer, Cham, 2015.

[34] M. Mareli and B. Twala, "An adaptive Cuckoo search algorithm for optimisation," *Applied computing and informatics*, vol. 14, no. 2, pp. 107–115, 2018.

[35] H. Wang, W. Wang, X. Zhou et al., "Firefly algorithm with neighborhood attraction," *Firefly algorithm with neighborhood attraction*, Information Sciences, vol. 382-383, , pp. 374–387, 2017.

[36] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications*, vol. 72, pp. 221–230, Apr. 2017.

[37] S. Xingjian, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolution LSTM network: a machine learning approach for precipitation nowcasting," in *Proceedings of the Advances in Neural Information Processing Systems*pp. 802–810, Montreal, QC, Canada, 2015.

Oxidation Communications 45, No 4, 701–713 (2022)

Biological and biochemical oxidation processes; medical image processing; neural networks and fuzzy logic techniques

IMPLEMENTATION OF FEATURE SELECTION AND ENSEMBLE DEEP LEARNING CLASSIFIER FOR PARKINSON'S DISEASE

B. SABEENA*, S. SIVAKUMARI

Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering, 641 043 Coimbatore, Tamil Nadu, India
E-mail: *sabeenar09@gmail.com*

ABSTRACT

Parkinson's diseases are chronic neuro-degenerative conditions that impact humans in their day to day lives. Diagnosis and monitoring of these conditions, based on limited physical symptoms, are painstaking evaluations for medical professionals and clinicians may miss early prodromal phases. Though many data mining techniques for automated assessments of Parkinson's diseases have recently been presented, their performances get reduced due to dataset's irrelevant features. Ensemble feature selections have more benefits than Single feature selection algorithms as they address drawbacks by mixing different models and improve outcomes of Machine learning techniques. This work uses Optimisation-based ensemble feature selections including Fuzzy monarch butterfly optimisation algorithms, Levy flight cuckoo search algorithms, and Adaptive firefly algorithms for selection of features based on their correlations. On selection of features, Ensemble deep learning classifiers classify Parkinson's diseases datasets. Ensembe deep learing include Fuzzy convolution bi-directional long short-term memories, Contractive auto-encoders, and Sparse auto-encoders. Sparse auto-encoders are robust variant of standards of auto encoders which learn representations with reduced sensitiveness to small variations of data. They are used to train classifiers using neural networks (NNs) for identifying Parkinson's diseases from datasets. Stacked generalisation is used to combine the results of DL classifiers. When compared to a single model, Ensembe deep learing techniques offer improved predictive performances. The datasets used for this study were obtained from machine learning repositories of University of California-Irvine. The performance of the classifiers was measured using accuracies, *F*-measures, MCCs (Matthews Correlation Coefficients), and errors.

Keywords: Parkinson's Diseases (PDs), Data mining techniques (DMTs), Optimisation-based ensemble feature selection (OBEFS), Levy flight cuckoo search algorithms (LFCSAs), Ensemble deep learning (EDL) classifier.

* For correspondence.

AIMS AND BACKGROUND

PDs are degenerative neurological illnesses characterised by low dopamine levels in the brain[1]. Inadequate, intermittent symptom monitoring, infrequent access to care, and few interactions with healthcare experts restrict PDs patient care, resulting in poor medical decision making and suboptimal patient health-related outcomes. PDs are the second most common neurological illnesses behind Alzheimer's[2,3]. Tremors, stiffness, bradykinesia, and postural instabilities are the four main signs of PDs (Ref. 4). These symptoms are persistent and degenerative and worsen over a period of time, however, they appear in varying degrees and combinations based on individuals. PDs may include both motor and non-motor symptoms that can make it difficult for affected patients to live normally[5]. Over 90% of patients with PDs have vocal impairments including dysphonic (difficulty in producing sounds vocally). Recent studies have also discovered relationships between risk alleles counts and health and the probability of developing PDs (Ref. 6). Around 1–2% of adults over the age of 60 globally have been impacted by the condition.

Systems detecting PDs use many sensors for determining symptom's severities. A major symptom that is evident in patients with PDs is vocal presentations which they suffer in early stages of diseases. Thus, detecting impairments in vocal presentations are primary to investigations of PDs (Ref. 7). The patient's vocal recordings have been exploited by algorithms for generating relevant features and input into many systems that learnt from these characteristics and used it for classifications. Machine learning techniques (MLTs) including Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) (Ref. 8) have been common in studies which classified PDs in addition to Random forests (RFs) and K-nearest neighbours (KNNs) which are simple and convenient. The quality of the data attributes are intricately related to success of aforementioned algorithms. Although manually identifying relevant features to characterise inherent attributes of speech (audio) data is complex, data's latent properties can be determined automatically using a Deep learning techniques (DLTs) (Ref. 9).

FSAs (Feature selection algorithms) play crucial roles in reducing training time by their reductions of features and by their selections of relevant features and eliminations of redundant features with the aim of improving classifier performances. There are three forms of FSAs (selecting subsets of features without assessments) and hybrid techniques. They include wrappers which are search algorithms to find and estimate relevant subsets of characteristics and filters which are combinations of previous methods. Though there are several FSAs, there are no tools or solutions that can objectively determine which algorithms work best with given datasets. Hence, trial-and-error strategies are used in various inquiries[10].

In this research, OBEFSs method has been suggested to choose features based on the PD datasets. The OBEFS algorithm, which seeks to merge numerous feature selection approaches via Fuzzy Monarch Butterfly Optimisation Algorithms (FMBOAs), LFCSAs, and AFAs, is introduced. The optimal features selected through OBEFS are used to train an EDL (Ensemble deep learning) classifier. Stacking generalisation is used to integrate EDL classifiers such as Fuzzy convolution bi-directional long short-term memories (FCBi-LSTMs), Contractive Auto-encoders (CAEs), and Sparse auto-encoders (SAEs). Proposed method is trained with a dataset taken from machine learning repositories of University of California-Irvine (UCI) and their performances. Evaluation metrics such as *F*-measure, Matthews correlation coefficients (MCCs), accuracies, and errors are used for the assessment of classification.

EXPERIMENTAL

This work suggests FSAs and classification scheme for approach for identifying PDs in patients. KPCAs (Kernel-based principal component analyses) reduce dimensionalities followed by OBEFSs executions which include FMBOAs, LFCSAs, and AFAs. EDL classifiers such as FCBi-LSTMs, CAEs , and SAEs were used in classifications. Finally all classifiers results were combined using stacked generalisations and the resultant outputs were evaluated with performance metrics like *F*-measures, MCCs, accuracies, and errors. Figure 1 depicts the overall flow of this work's suggested scheme.

Parkinson's disease dataset. This work used PDs dataset from machine learning repositories of UCI. The dataset used in the study was obtained from Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University and included 188 patients with PDs (107-males, 81-females) and 64 non-diseased individuals (23-males and 41-females). The patient's age ranges were 33–87. The ages of healthy individuals ranged between 41 and 82 years. The voices were recorded at 44.1 kHz and following doctor's inspections, three duplicates of the vowel /a/ letter of persons were obtained.

Feature extraction. Features including Baseline features, Time frequency Features, MFCCs (Mel-Frequency Cepstral Coefficients), WTs (Wavelet Transforms), TQWTs (Tunable Q-factor WTs), and Vocal fold features were retrieved from the dataset.

(i) Baseline characteristics: Patients affected with PDs experience impaired speech even during early stages of the disease's onset and hence characteristics of speech can be effectively utilised to evaluate PDs and track its progression for medical therapies. Examples of baseline features can be: Jitter and glow-based features, fundamental frequency parameters (#5), harmonicity parameters (#2), RPDEs (Recurrence time density entropies) (#1), DFAs (Detrended fluctuation analyses) (#1), and PPEs (Pitch period entropies) (#1) are commonly used speech characteristics in Parkinson's disease studies.

(ii) Time frequency features: The features like Intensity parameters (#3), Formant frequencies (#4) and Bandwidth (#4).

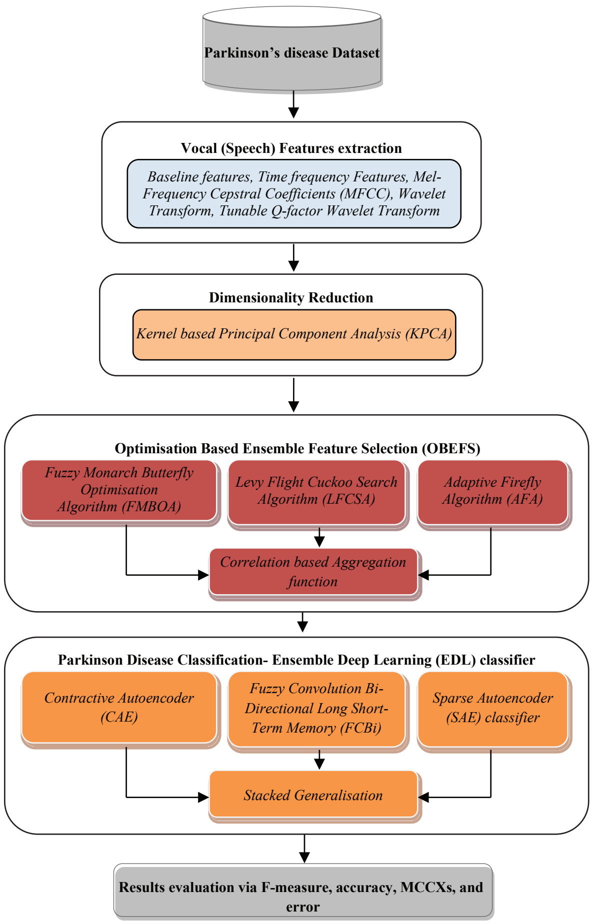


Fig. 1. Overall flow of the suggested scheme

(iii) MFCCs: MFCCs-based extraction methods use triangular overlapping filter banks to combine cepstral analysis with spectral domain partitions. The data contained 84 MFCCs-related parameters generated using mean and standard deviations of the first 13 MFCCs, as well as the signal's log energy's 1st and 2nd, apart from the vocal folds, to identify PD effects in the vocal tract (#84).

(iv) WTs: WTs are common approaches for making decisions about signals in general, especially ones with minor geographical variances. This approach produces 182 WT-based characteristics for both the approximation and detailed coefficients, including energy, Shannon's and log energy entropy, and Teager-Kaiser energy.

(v) TQWTs: TQWTs use three adjustable parameters (Q (Q-factors), r (redundancies), and J (count of levels)) to improve the quality of signal conversion based on signal behaviour. Using this dataset, many tests yielded 432 TQWT-related variables.

(vi) Vocal fold characteristics: Vibrations of voices in features were examined for noises. The data were used to extract features like GQs (Glottis quotients) (#3), GNEs (Glottal to noise excitations) (#6), VEFRs (Vocal fold excitation ratios) (#7), and EMDs (Empirical mode decompositions) features.

(vii) Concat characteristics: Concat features combine baseline, vocal fold and time frequency features.

Reducing dimensionalities with KPCAs. KPCAs reduce dimensionalities, i.e. they reduce linear dimensions of the high dimensional sound waves for examining variances in the case of PDs. Assume that PDs dataset is labelled i, $i = 1, ..., N$, with each i being a D-dimensional sound recording characteristics vector. The sound recordings' dimensionality reduction feature vector is found using this vector. OBEFSs are utilised for feature selection using dimensionality reduced feature vectors.

FMBOAs. FMBOAs are methods for choosing feature sets. FMBOAs select characteristics for samples based on the degree of existence effects. MBOs are based on the migratory behaviours. FMBOAs have been exploited for good results in classifications when employed without modifications. This implies that they manage to balance their searches both global and local. Butterflies relate and distribute information locally within the swarm and thus assisting in enhanced information for systems where the search operations and sharing of information are algorithmically executed using operators for migrations and butterfly adjustments.

LFCSAs. CSAs (Cuckoo search algorithms) are based on cuckoo species which use suitable hosts to optimise selection of attributes from datasets in order to nurture their brood. Some cuckoo species employ a suitable host to optimise the selection of features from a dataset in order to nurse their brood. CSAs are used to reduce the risk of egg losses (irrelevant features) to other species without avoiding parental commitments in raising their progeny. The final qualities are determined by putting eggs (features) in several nests. LFCSAs start with a population of N host nests. A cuckoo's egg (feature), say i, is picked at random at each iteration $\underline{t}$, and new feature solutions f_i^{t+1} are generated. The Lévy flights are a sort of random walks where steps are characterised in terms of step-lengths and following certain probability distributions. The step's orientations must be isotropic and random. Probabilities prb_a are used to compute replacement rates in a stochastic manner which are tuned for better performances. The best solutions (features) achieved so far are saved as feature vectors f_{best}, and all current solutions (feature selection solutions) are ordered according to their fitnesses (accuracies) at iterations. The procedure is repeated until the specified stopping threshold is reached.

AFAs. The idealised behaviour of firefly flashing is the basis for FAs. Fireflies steps need to placed far away from optimal answers as possible. For the optimal selection of attributes from datasets, fireflies are employed and global and local searches are balanced. As a result, firefly's strides also need to consider past and current location's data. Historical data of fireflies which contain optimal values of past two iterations are taken into account in this work. The distances between current fitness values and population's best fitness values determine further steps of fireflies in iterations. The steps may vary with iterations, and steps of fireflies are also altered in iterations.

Correlation function. Correlation coefficient matrices from features were computed for selecting outputs of ensemble's feature selection and were determined using equation (1):

$$\text{correlation coefficient} = \frac{N \sum xy - \sum x \sum y}{((N \sum x^2 - (\sum x)^2)(N \sum y^2 - (\sum y)^2))^{1/2}}, \quad (1)$$

where x, y represent examined attribute values while N stands for instance counts. Feature sets chosen based on correlations by ensembles become inputs for classifiers.

Classifications of PDs via Ensemble deep learning (*EDL*) *classifier*. In this work, PDs are classified using an EDL classifier. Ensemble learning can considerably boost the learning system's generalisation capabilities. EDL, a machine-learning approach that uses numerous classifiers such as FCBi-LSTMs, CAEs, and SAEs to construct an ensemble learner using stacked generalisation to enable greater generalisation of learning systems, has shown remarkable success in classifications of PDs. Deep EDL combines the benefits of three deep learning models as well as ensemble learning to provide a model with improved generalisation performance. The EDL classifier learns numerous base classifiers and aggregates their results using specified criteria. The rule used to aggregate the outputs defines an ensemble's effective performance.

FCBi-LSTMs. The FCBi-LSTMs are used for classifications based on characteristics chosen from PD datasets. Convolution and pooling layers make up these networks. They carry out convolutions then pooling where outputs are provided as inputs for subsequent convolution layers. Convolution layers using filters divide features into multiple matrices of sizes ($Nm + 1$) for better representations. These resultant matrices are reduced in dimensions by CNN's pooling layers but manage to maintain links between features. These averaging layers which use preceding convolved inputs, send their outputs as inputs to subsequent layers. Bi-LSTMs, use final convolution layer's outputs as intermediary variables and hence perform better than uni-directional LSTMs and retain more structural information. The final outputs of Bi-LSTMs are processed by two convolution layers to obtain diagnosis of PDs. MFB (Multi-modal factorised bilinear pooling) is a technique for combining information from CNNs with Bi-LSTM's obtained features.

SAEs. The SAEs classifier neurons labelled as (+1) are bias units introduced to the FFNNs (feed-forward NNs) through cost functions. This phase drives AEs more correctly to replicate inputs x without over fits. Figure 2 depicts an overview of the model. The sparse auto encoder bottlenecks are utilised as input vectors to DNNs.

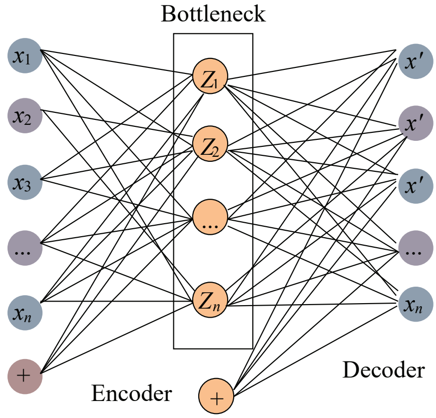


Fig. 2. Sparse model of SAEs

SAEs are proposed in which cost functions are made of 3 parts and are detailed below. Assuming dataset has N training instances $(x_1, x_2, ..., x_n)$, where x_i stands for i-th input. SAEs learn to reconstruct inputs x_i using cost functions $h_{\mathrm{W,b}}(x_i)$ nearing x_i where MSEs, weight decays, and sparsity are included in the function. The cost functions for N training samples MSEs and weight decays are defined in the following equation:

$$J_{\mathrm{sparse}}(W, b) = (1/N) \sum_{i=1}^{N} (1/2)\|h_{\mathrm{W,b}}(x^i) - x^i\| + (\lambda/2) \sum_{l=1}^{n_{l-1}} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{\ l})^2. \tag{2}$$

Weight decays specified by the equation above avoids over fits of data, since small values of λ can lead to data over fits, while its larger values can result in data under fits. Sparsity, the third component of the cost functions, helps in activating hidden layers of AEs and prevents data over fits. It has the ability to minimise the number of hidden layer areas evaluated. The average active value of the hidden layer was calculated using the following equation, where a denotes the activation function, which is rectifier (ReLU):

$$\hat{p}_j = (1/N) \textstyle\sum_{i=1}^{N} (a_j^{\ 2}(x^i)). \tag{3}$$

Sparsity is computed for getting $\hat{p}_j$ nearer to p (sparsity parameter), helps in deviations from p and results in activating or de-activating hidden layer's neurons. It can be defined using the Kullback-Leibler divergences as depicted in equation (4):

$$\textstyle\sum_{j=1}^{sl} KL(p\|\hat{p}_j) = \sum_{j=1}^{sl} (p \log (p/\hat{p}_j) + (1 - p) \log ((1 - p)/(1 - \hat{p}_j)). \tag{4}$$

SAEs cost functions add all three component results and is depicted as equation (5):

$$J_{\mathrm{sparse}}(W, b) = (1/N) \sum_{i=1}^{N} (1/2)\|h_{W,b}(x^i) - x^i\| + (\lambda/2) \sum_{l=1}^{n_{l-1}} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{\ l})^2 + \beta \sum_{j=1}^{s_l} KL(p\|\hat{}p_j), \tag{5}$$

where β represents sparse penalties. The deep NNs classifier is trained using the bottleneck of the SAEs as inputs, and the SAEs are trained to minimise their cost function as mentioned above. The SAEs and the classifier were both trained at the same time, resulting in improved feature extraction while the classifier's choice was optimised. The training procedure takes 30 iterations, using 8 instances per batch. The sparsity parameter p was set to 0.05 while weight decay λ was set to 0.0001 and sparse penalty term to 2. DNN fine-tuning on the final 10 iterations to change classifier parameters and reduce the softmax cost function while SAE parameters remain fixed. The parameters are updated using the Adam optimiser depending on the calculated gradients.

CAEs. Contractive AEs transform learned representations are robust towards small changes around the training examples. It achieves that by using different penalty term imposed to the representation. The loss function (l_2) is used for the reconstruction term. AEs use the encoder function f to map inputs x to internal representations (or codes) $f(x)$. The decoder functions g are then used to map $f(x)$ back to their input spaces. Reconstruction functions r made up of the functions f and g, i.e. $r(x) = g(f(x))$, and reconstruction loss functions l penalises errors, with $r(x)$ considered as forecasts of x. CAEs are regularised AEs that learn to minimise the regularised reconstruction errors and are depicted below:

$$L_{\mathrm{CAE}} = E[l(x, r(x)) + \lambda\|\partial f(x)/\partial x\|_{\mathrm{F}}^2], \tag{6}$$

where $r(x) = g(f(x))$ and $\|A\|_{\mathrm{F}}^2$ is the sum of the squares of the elements of A. Both the squared loss:

$$l(x, r) = \|x - r\|^2 \tag{7}$$

and the cross-entropy loss:

$$l(x, r) = -x \log r - (1 - x) \log (1 - r). \tag{8}$$

Because of the simpler mathematical method it enables, concentrate your investigation on the squared loss. It is worth noting that minimising the CAEs criteria is very dependent on the parameterisation of f and g, particularly the linked weights restriction imposed by equation (9):

$$f(x) = \mathrm{sigmoid}(Wx + b) \tag{9}$$

$$g(h) = \mathrm{sigmoid}(W^T h + c). \tag{10}$$

Because of the linked weights, the foregoing regularising term compels f (as well as g) to be contractive, i.e. to have singular values < 1. Larger values of λ produce greater contractions (smaller singular values) where it has least impacts on reconstruction errors, i.e. in local directions with little or no variability of data.

Stacked generalisation. Stacked generalisation deduces generalised biases in relation to given learning sets. Cross validation data and least squares with non-negativity constraints were used to determine the best weights of combination in regression to generate a good linear combination of the base learners. Consider the linear combination of the base learners' predictions $f_1, f_2, ..., f_m$ given by equation (11):

$$f_{\text{stacking}}(x) = \sum_{j=1}^{m} w_j f_j(x), \tag{11}$$

where w is the optimal weight vector learned by the meta learner.

RESULTS AND DISCUSSION

The findings of the present work's experimental concerning the suggested EDL classifier and its comparisons with other approaches including FCLSTM-CNNs (Fuzzy Convolution Long Short-Term Memories) based CNNs, FCBi-LSTMs, and CNNs are detailed in this section. MATLAB R2016a was used to conduct the studies on recognising PDs and classifying them. The following system requirements were followed during implementation: Intel(R) Core™i3-4160T CPU@3.10 GHz 3.09 GHz processor, 4.00 GB RAM, Windows 8.1 pro, 64 bit operating system, operation system, and 1 TB hard disk.

Evaluation metrics. To analyse prediction performances of classifiers, evaluation metrics are required. Confusion matrix in Table 1 represents the counts of properly and wrongly categorised occurrences for classes based on binary classifications. The symbols tp, fp, fn, and tn in the confusion matrix represent true positive (tp), false positive (fp), false negative (fn), and true negative (tn), respectively. Precisions, recalls, F-measures, accuracies, and error rates were computed using formulae given in equations (12)–(16):

$$\text{precision} = tp/(tp + fp) \tag{12}$$

$$\text{recall} = tp/(tp + fp) \tag{13}$$

$$F\text{-measure} = (2 \times \text{precision} \times \text{recall})/(\text{precision} + \text{recall}) \tag{14}$$

$$\text{accuracy} = (tp + tn)/(tp + tn + fp + fn) \tag{15}$$

$$\text{error rate} = 100 - \text{accuracy}. \tag{16}$$

Another statistics for evaluating the quality of binary classifications is the MCCs. MCCs consider tp, fp, fn, and tn counts and are recognised as balanced measures that can be applied even when class distributions are imbalanced. MCCs are correlation coefficients between actual and projected occurrences that range in the interval [–1, +1] where +1 represents flawless predictions and –1 shows differences between forecasts and actual in labels.

Table 1. Confusion matrix for two-class classification

Actual/Predicted as	Positive	Negative
Positive	*tp*	*fn*
Negative	*fp*	*tn*

Results comparison. Experimentations using 4 features and classifiers were evaluated using the metrics of accuracies, errors, *F*-measures, and MCCs. The combination of MFCC + Wavelet + Concat features with a CNNs classifier yields an accuracy rating of 94.1752%. Despite the fact that the accuracy percentage of the MFCC + Wavelet + Concat combination is 95.1557% for the FCLSTM-C FCBi-LSTM classifier gives the accuracy results of 96.6381, 98.0244, 97.3457 and 98.7720% for TQWT + MFCC + Wavelet, TQWT + Wavelet + Concat, TQWT + MFCC + Concat, and MFCC + Wavelet + Concat, respectively. Proposed EDL classifier with MFCC + Wavelet + Concat combination achieves the highest performance with an accuracy rate of 99.903%, *F*-measure rate of 99.633%, and 72.431% for MCCs (Table 2).

Table 2. Results of classifiers with triple feature (KPCA + OBEFS)

CNNs classifier (%)				
Feature combination	*F*-measure	Accuracy	Error rate	MCCs
F1-TQWT + MFCC + Wavelet	85.8697	87.5696	12.4304	57.2007
F2-TQWT + MFCC + Concat	90.2315	91.9315	8.0684	61.4600
F3-TQWT + Wavelet + Concat	86.3695	88.0694	11.9306	63.3994
F4-MFCC + Wavelet + Concat	92.4752	94.1752	5.8248	64.5384
FCLSTM-CNNs classifier (%)				
F1-TQWT + MFCC + Wavelet	94.2258	93.0470	6.9530	67.6669
F2-TQWT + MFCC + Concat	91.5250	93.0854	6.9146	67.7060
F3-TQWT + Wavelet + Concat	93.4200	93.1261	6.8739	65.4457
F4-MFCC + Wavelet + Concat	91.6921	95.1557	4.8443	67.2960
FCBi-LSTM classifier (%)				
F1-TQWT + MFCC + Wavelet	98.3100	96.6381	3.3619	74.300
F2-TQWT + MFCC + Concat	96.5900	98.0244	1.9756	72.300
F3-TQWT + Wavelet + Concat	97.5200	97.3457	2.6543	70.300
F4-MFCC + Wavelet + Concat	98.5010	98.7720	1.2280	71.400
EDL classifier (%)				
F1-TQWT + MFCC + Wavelet	99.449	97.771	2.2299	75.432
F2-TQWT + MFCC + Concat	97.722	99.156	0.8436	73.412
F3-TQWT + Wavelet + Concat	98.652	98.378	1.6222	71.013
F4-MFCC + Wavelet + Concat	99.633	99.903	0.0966	72.431

Figures 3–6 show *F*-measures, accuracies, MCCs, and errors using different feature sets. When compared to other feature combination sets, TQWT + MFCC + Wavelet combination of feature set yields greater results of 99.449, 97.77, 75.432, and 2.2299% for *F*-measure, accuracy, MCCs, and error, respectively. The suggested classifier using the first feature level combination achieves a higher *F*-measure of 99.449%, whereas other approaches such as CNNs, FCLSTM-CNNs, and FCBi-LSTM achieve *F*-measures of 85.8697, 94.2258, and 98.3100%, respectively.

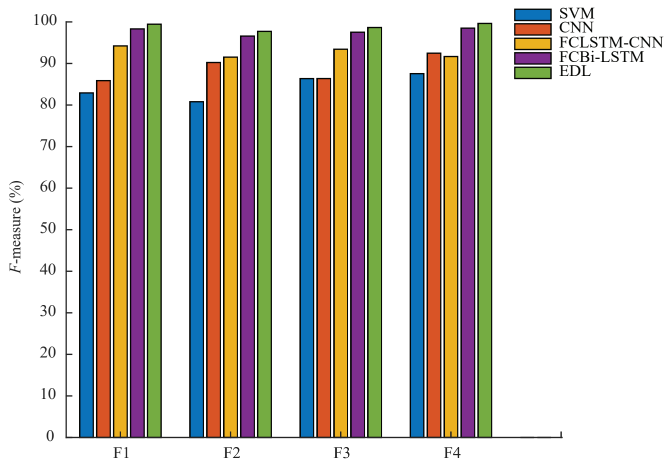


Fig. 3. *F*-measure comparison of feature level combination versus classifiers

Figure 4 shows that the *x*-axis results are measured using a feature-level combination of classifiers, and the *y*-axis results are shown as accuracy. The final feature level combination proposed for the classifier achieves 99.9030% accuracy, whereas other approaches such as CNNs, FCLSTM-CNNs, and FCBi-LSTM achieve 94.1752, 95.1557, and 98.7720% accuracy, respectively.

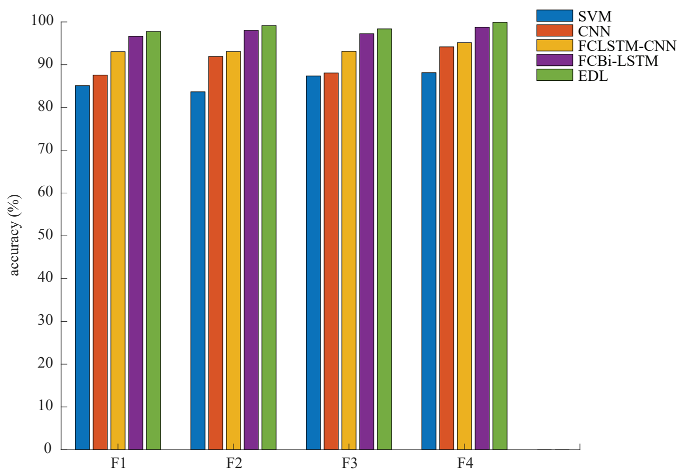


Fig. 4. Accuracy comparison of feature level combination versus classifiers

Figure 5 compares the error rates of classifiers with four different feature level combinations. It shows that the proposed classifier with the final feature level combination has a lower error rate of 0.0966%, whereas CNNs, FCLSTM-CNNs, and FCBi-LSTM approaches have larger error rates of 5.8248, 4.8443, and 1.2280%, respectively.

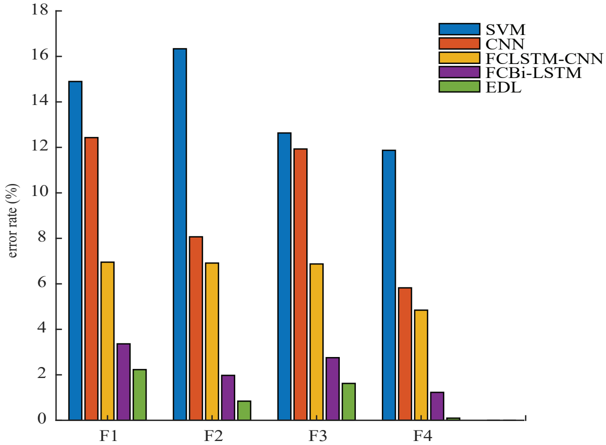


Fig. 5. Error rate comparison of feature level combination versus classifiers

Figure 6 displays MCCs results comparison of first feature level combination produces greater results of 75.432% for proposed algorithm, 57.2007, 67.6669, and 74.300 % for CNNs, FCLSTM-CNNs, and FCBi-LSTM approaches classifiers, accordingly.

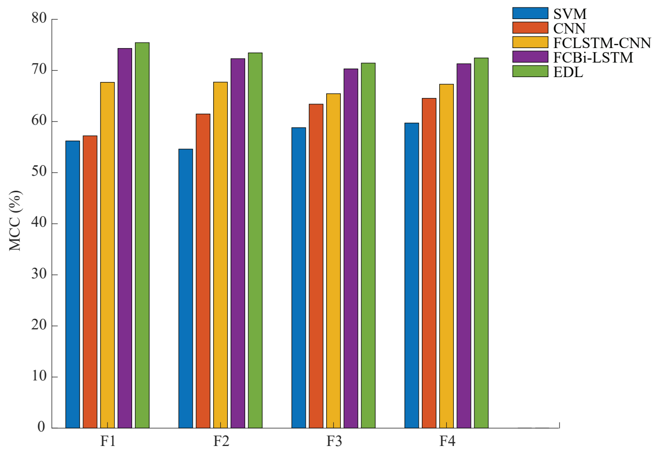


Fig. 6. MCCS comparison of feature level combination versus classifiers

CONCLUSIONS

In this study, an ensemble learner known as OBEFSs is proposed along with Ensemble Deep Learning (EDL) classifier that combines single approaches. This is done using algorithms such as FMBOAs, LFCSAs, and AFAs. To carry out OBEFS, the correlation function is used to choose the best features from the three feature subsets. After that,

ENSEMBLE FEATURE SELECTION AND ENSEMBLE DEEP LEARNING (EDL) CLASSIFIER FOR PARKINSON'S DISEASE

B. SABEENA, S. SIVAKUMARI,

Research Scholar, Department of Computer Science and Engineering, Avinashilingam Institute forHome Science and Education for Women, School of Engineering

Professor, Department of Computer Science and Engineering, Avinashilingam Institute for Home Science and Education for Women, School of Engineering.

E-mail: sabeenar09@gmail.com

ABSTRACT

PDs (Parkinson's diseases) are chronic neuro-degenerative conditions that impact humans in their day to day lives. Diagnosis and monitoring of these conditions based on limited physical symptoms are painstaking evaluations for medical professionals and clinicians may miss early prodromal phases. Though many DMTs (data mining techniques) for automated assessments of PDs have recently been presented, their performances get reduced due to dataset's irrelevant features. EFSs (Ensemble Feature Selections) have more benefits than single FSAs (Feature Selection algorithms) as they address drawbacks by mixing different models and improve outcomes of MLTs (machine learning techniques). This work uses OBEFSs (Optimization Based Ensemble Feature Selections) including FMBOAS (Fuzzy Monarch Butterfly Optimization Algorithms), LFCSAs (Levy Flight Cuckoo Search Algorithms), and AFAs (Adaptive Firefly Algorithms) for selection of features based on their correlations. On selection of features, EDL (Ensemble Deep Learning) classifiers classify PD datasets. EDLs include FCBi-LSTMs (Fuzzy Convolution Bi-Directional Long Short-Term Memories), CAEs (Contractive Auto-encoders), and SAEs (Sparse Auto-encoders). CAEs are robust variant of standards of auto encoders which learn representations with reduced sensitiveness to small variations of data. SAEs are used to train classifiers using NNs (neural networks) for identifying PDs from datasets. Stacked generalization is used to combine the results of DL classifiers. When compared to a single model, EDL techniques offer improved predictive performances. The datasets used for this study were obtain ed from machine learning repositories of UCI (University of California-Irvine). The performance of the classifiers were measured using accuracies, F-measures, MCCs (Matthews Correlation Coefficients), and errors.

Index Terms: *Parkinson's Disease (PDs), Optimization Based Ensemble Feature Selection (OBEFS), Levy Flight Cuckoo Search Algorithm (LFCSA), and Adaptive Firefly Algorithm (AFA), Ensemble Deep Learning (EDL) Classifier, Fuzzy Convolution Bi-Directional Long Short-Term Memory (FCBi-LSTM), Contractive Autoencoder (CAE), and Sparse Autoencoder (SAE).*

1. INTRODUCTION

PDs are degenerative neurological illnesses characterised by low dopamine levels in the brain [1]. Inadequate, intermittent symptom monitoring, infrequent access to care, and few interactions with healthcare experts restrict PDs patient care, resulting in poor medical decision making and suboptimal patient health-related outcomes. PDs are the second most common neurological illnesses behind Alzheimer's [2-3]. Tremors, stiffness, bradykinesia, and postural instabilities are the four main signs of PDs [4]. These symptoms are persistent and degenerative and worsen over a period of time, however they appear in varying degrees and combinations based on individuals. PDs may include both motor and non-motor symptoms that can make it difficult for affected patients to live normally [5]. Over ninety percent of patients with PDs have vocal impairments including dysphonic (difficulty in producing sounds vocally). Recent studies have also discovered relationships between risk alleles counts and health and the probability of developing PDs [6]. Around 1–2 percent of adults over the age of 60 globally have been impacted by the condition.

Systems detecting PDs use many sensors for determining symptom's severities. A major symptom that is evident in patients with PDs is vocal presentations which they suffer in early stages of diseases. Thus, detecting impairments in vocal presentations are primary to investigations of PDs [7–8]. The patient's vocal recordings have been exploited by algorithms for generating relevant features and input into many systems that learnt from these characteristics and used it for classifications. MLTs including ANNs (Artificial Neural Networks) and SVMs (Support Vector Machines) [9] have been common in studies which classified PDs in addition to RFs (Random Forests) [10] and KNNs (K-Nearest Neighbors) [11] which are simple and convenient. The quality of the data attributes are intricately related to success of aforementioned algorithms. Although manually identifying relevant features to characterize inherent attributes of speech (audio) data is complex, data's latent properties can be determined automatically using a DLTs (deep learning techniques).

FSAs play crucial roles in reducing training time by their reductions of features and by their selections of relevant features and eliminations of redundant features with the aim of improving classifier performances. There are three forms of FSAs (selecting subsets of features without assessments) and hybrid techniques. They include wrappers which are search algorithms to find and estimate relevant subsets of characteristics and filters which are combinations of previous methods. Though there are several FSAs, there are no tools or solutions that can objectively determine which algorithms work best with given datasets. Hence, trial-and-error strategies are used in various inquiries. Studies explored variety of FSAs with one or more classifiers before selecting ones that performed best in tests [11,12,13]. Alternatively, ensemble learning algorithms have been described for picking features based on consensus or aggregates of many FSAs [14,15].

In ensemble learning, SIs (swarm intelligences) are decentralised techniques that are self-organizing with collective behaviours . They usually comprise of basic agent groups that interact with one another and their environments at local levels. There are no centralised control structures to decide single agent's behaviours, and agents follow basic principles. On the other hand, interactions of these entities result in emergence of "intelligent" global behaviours . SIs can also use indirect communications to exchange and coordinate information. The increase in communication overhead is minor as the number of people increases. As a result, it is also scalable. Recent works have employed rigorous FSAs based on EFSs where results suggested that the study's proposed approach had significant potential in selecting features from datasets with more features and lesser sample counts. Many research on EFSs have recently been undertaken; some use classifiers, while others do not. Population-based optimization techniques such as ACOs (Ant Colony Optimizations) . GAs (Genetic Algorithms), SAs (Simulated Annealing) . TSs (taboo searches) [20], and PSOs (Particle Swarm Optimizations) have recently been employed as FSAs. Hybrid search strategies that merge wrappers and filters have also been used.

This work's OBEFSs framework aims to assist in the creation of EFSs mechanisms that combine benefits of several FSAs while avoiding biases and compensating for downsides. SIs have been extensively used in feature selections because of advantages namely, their combinations with MLTs produce outstanding results. DNNs (Deep Neural Networks) with their hierarchical layers produce deep abstract representations, which are used as inputs in many MLTs. These performances have prompted researchers to apply DNNs in classification of PDs. DNNs is a potential classifier for classifications of PDs since it can represent intricate and non-linear relationships from data.

Second, the class of provided data samples predicted in classifications are a type of supervised learning. Ensemble learning evaluates a variety of approaches instead of single classification algorithms, and final results are generated by merging outputs of classifiers. Main objective of ensemble classifiers is to combine advantages of numerous classifiers and integrate their outputs, such that Base classifiers are individual classifiers. Ensemble classifiers have two main problems: (1) selecting basic classifiers and (2) aggregating outputs of base classifiers . It is critical to ensure that basic classifiers are sufficiently diversified while forming successful ensembles . Ensemble learning is a powerful approach that combines numerous learning algorithms to increase overall prediction accuracy and may exceed any single smart classifier.

In this research, OBEFSs method has been suggested to choose features based on the PD datasets. The OBEFS algorithm, which seeks to merge numerous feature selection approaches via FMBOAs, LFCSAs, and AFAs, is introduced. The optimal features selected through OBEFS are used to train an EDL (Ensemble Deep Learning) classifier. Stacking generalisation is used to integrate EDL classifiers such as FCBi-LSTMs, CAEs, and SAEs. Proposed method is trained with a dataset taken from machine learning repositories of UCI and their performances. Evaluation metrics such as F-Measure, MCCs, accuracies, and errors are used for the assessment of classification.

2. LITERATURE REVIEW

MLTs for recognizing PDs. Their scheme used hybrid FSAs by combining ACOs and Relief networks. The feature outputs FSAs were used for classifications by SVMs for maximum classification accuracies. The study's schemes were evaluated using K-fold cross validations for justifying hyper

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

parameters. Experimental results of the scheme from real world PD datasets, showed that their proposed system outperformed baseline techniques in recognising PDs from specified attributes. The performances suggested that their approach was highly recommendable for identification of PDs.

a new class of OCSAs (Optimized version of Crow Search Algorithms). Their recommended OCSAs could be used to predict PDs and help people receive appropriate treatments at early stages. The performance of OCSAs were evaluated for 20 benchmark datasets and the results compared to original CCSAs (Chaotic Crow Search Algorithms). The proposed nature-inspired algorithm discovered ideal subsets of characteristics, maximized accuracy while minimizing selected features according to their findings.

an improved FKNNs (Fuzzy KNNs) approach based on voice measures for the early diagnosis of PDs. The suggested method called CBFO-FKNNs used evolutionary instance based learning strategies where CBFOs (Chaotic Bacterial Foraging Optimizations) with Gauss mutations and FKNNs were used. CBFOs examined parameters before tuning them for use by FKNNs. The scheme's obtained values of accuracies, sensitivities, specificities, and AUC (Area Under the Curves) were compared with other methods using ten fold cross validations for PD datasets. The study's suggested scheme outperformed other FKNN models based (Five) on BFOs (Bacterial Foraging Optimizations), PSOs, GAs, FFOs (Fruit Fly Optimizations) and FAs (Firefly Algorithms) and MLTs (three) including SVMs, local learning feature selections based SVMs, and KELMs (kernel Extreme Learning Machines). The proposed approach provided work has a very strong possibility of bringing tremendous ease to physicians in making better clinical diagnosis decisions.

To choose the best features from the speech collection, suggested MAFTs (Multi-Agent Feature Filters). MAFTs aimed at selecting sets of characteristics to improve overall prediction model's performances and reduce over-fits which might occur due to the extreme reduction of features. Furthermore, this approach minimizes prediction complexities, expedites training, and develops robust training models. MAFTs were then combined with ten different MLTs to build a sophisticated voice-based detection model for PDs. To increase classification accuracies for detecting PDs, a Hybrid Model composed of binary CNNs (Convolution Neural Networks) and three FSAa namely, GAs, Adam optimizers, and mini-batch gradient descents was presented. Their results suggested that combination of MAFTs with hybrid modela greatly increased diagnostic outcomes for PDs.

MLTs for comparison of voice measurement features in patient datasets to determine if patients had PDs. The performances of basic classifiers like DTs (Decision Trees), LRs (Logistic Regressions), and KNNs were compared to Ensemble learning classifiers including Bagging, RFs, and Boosting. The accuracies of classifiers were found to be more accurate in predictions of sicknesses. Moreover, important traits needed for classifications were found based on feature's importance. The major purpose of the study was to detect dysphonia and distinguish healthy individuals from patients affected with PDs.

two frameworks based on CNNs for classifying PDs utilising sets of vocal (speech) characteristics. Before being integrated in layers, deep features from parallel branches were collected at the same time. The study's MLTs were trained on machine learning repositories of UCI while outcomes were verified using the metrics of LOPO CVs (Leave-One-Person-Out Cross Validations), accuracies, F-scores and , MCCs as the inputs had unequal class distributions.

Mutual Information Gains, extra trees, GAs, classifiers including NBs (Naive Bayes), KNNs, and RFs as FSAs. The study evaluated performances of various combinations on the speech dataset obtained from machine learning repositories of UCI. The study used SMOTEs (Synthetic Minority Oversampling Techniques) to handle class imbalances as the dataset was substantially uneven in instances. Their results from experiments were very useful.

model-based logics like LRs, KNNs, SVCs (Support Vector Classifiers), GBCs (Gradient Boost Classifiers) and RFs. The study evaluated reliability using five-fold cross-validations similar to curves of ROCs (Receiver Operating Characteristics) and confusion matrices. The study used majority voting, weighted averages, bagging, Ada boost, and GBCs in their ensembles. Their model also identified confusion matrices, five-fold cross-validations, precisions, recall rates, and F1 scores. The study's correlation matrices were also constructed to indicate if these characteristics were connected. Their findings implies that MLTs provided more reliable identifications of PDs in patients when compared to conventional techniques.

created a three-stage ensembles from DLTs for prognosis of PDs. To extract characteristics, the study employed DaTscan and clinical evaluations of motor complaints. Their ensembles of DNNs generated subsets of information gathered from patients four years after the onset of PDs to estimate PDs. Their proposed method was evaluated using MAPEs (Mean

Journal of Theoretical and Applied Information Technology
15th May 2023. Vol.101. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Absolute Percentage Errors), MAEs (Mean Absolute Errors), PCCs (Pearson's Correlation Coefficients) and biases between predicted and observed motor outcome scores. The study compared different data subsets as inputs to individual networks in evaluations.

dual layered stacking ensembles for accurately identifying PDs from healthy controls where multi-modal information were used. The first layer of their stacking ensemble architecture evaluated the advantages of four fundamental classifiers SVMs, RFs, KNNs, and ANNs while their second layer used LRs for categorizations. The performance of their proposed model was compared with standard ensembles.

a strategy that combined CARTs (Classification and Regression Trees) with ensembles. CARTs iteratively chose optimal training speech samples resulting in samples with high separability. Subsequently, these outputs were optimised by ensembles including RFs, SVMs, and ELMs. Their test data were classified using the trained ensembles. The study's recommended strategy was validated with other comparable methods using most recent datasets.

an ensemble-based technique for class label prediction based on voice frequency features for classifying sick and healthy individuals. Their scheme was divided into three stages namely data preparations, internal and final classifications. The findings of the study's tests showed that by using ensembles medical diagnostics were enhanced and provided comparative analyses of many MLTs. Internal classifier results were compared with sample's feature vectors.

a method using MENNs (Multi-Edit-Nearest-Neighbors). Initially, MENNs iteratively picked ideal training speech samples, resulting in samples with excellent separability. Subsequently, DNNEs (Decorrelated Neural Network Ensembles), learnt from acquired samples using ensembles. Finally, the taught ensembles were applied on test data for classifying PDs. The study compared their approach with currently used validation algorithms on latest public datasets.

DNNs as FSAs in an attempt to prove their efficacy by comparing performances of traditional DNNs with other integrated systems. The study developed EOFSCs (Ensembles of Optimal Features and Sample Dependant Base Classifiers) to capitalise on recent discoveries by studies. According to recent researches, distinct optimum models are developed for different forms of speech data that are sensitive to sample variations and subsets of attributes. Using the suggested integrated system, further consolidations of the findings were advised based on the development of EOFSCs. Basic classifiers show sensitiveness towards subsets of characteristics obtained from vowel phonations. This work's suggested EOFSCs use base classifier's for examining characteristics. The final forecasts of EOFSCs were evaluated using majority voting procedure. The results of their experiments suggested that combining FSAs with NNs improved performances of traditional methods. Moreover, integration of FSAs with h DNNs showed superior feature selection integrations with standard MLTs.

3. PROPOSED METHODOLOGY

This work suggest FSAs and classification scheme for approach for identifying PDs in patients. KPCAs (Kernel based Principal Component Analyses) reduce dimensionalities followed by OBEFSs executions which include FMBOAs, LFCSAs, and AFAs. EDL classifiers such as FCBi-LSTMs, CAEs , and SAEs are used in classifications. Finally all classifiers results are combined using Stacked Generalizations and the resultant outputs are evaluated with performance metrics like F-Measures, MCCs, accuracies, and errors. Figure 1 depicts the overall flow of this work's suggested scheme.

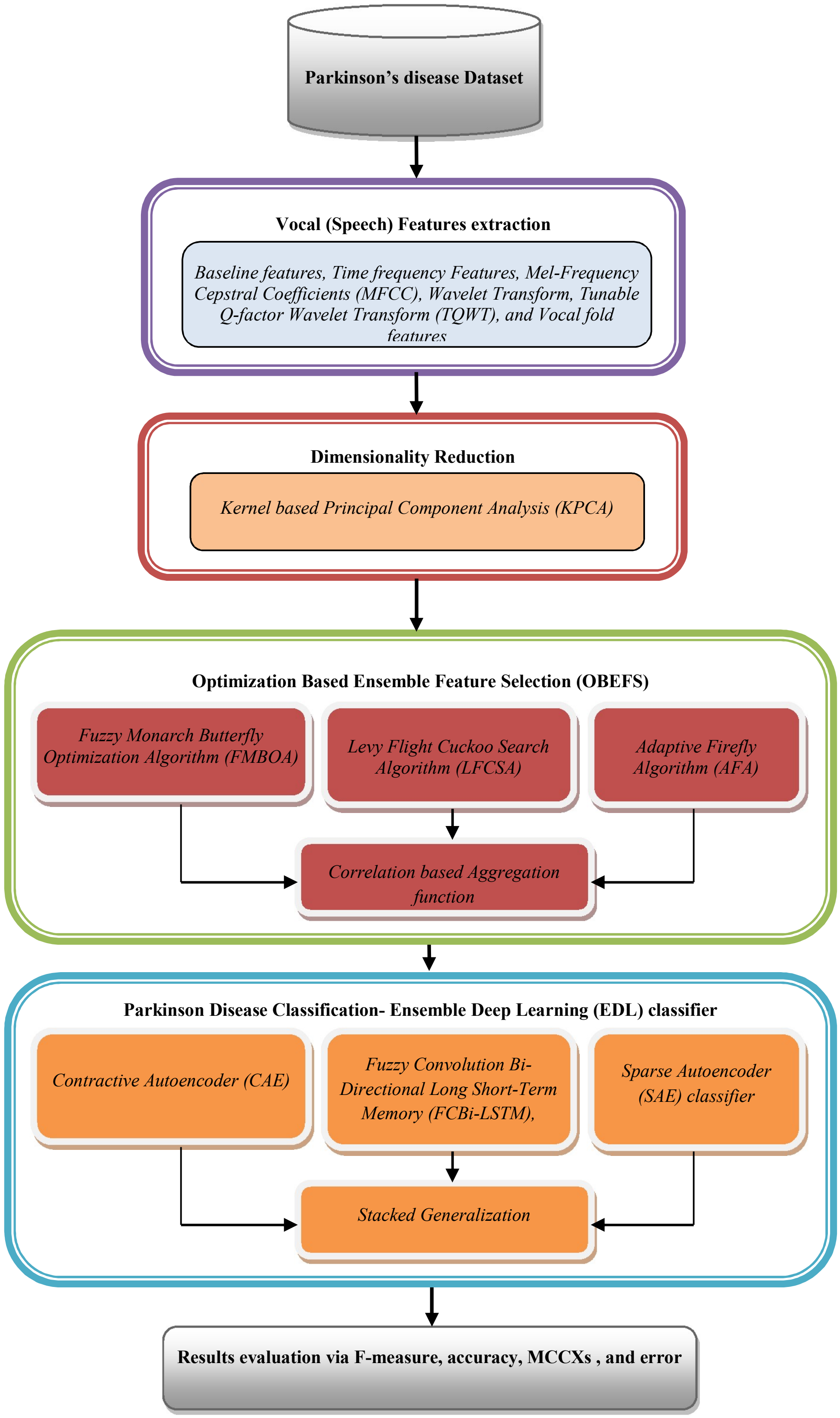


Fig. 1 - Overall Flow Of This Work's Suggested Scheme.

3.1. Parkinson's disease Dataset

This work used PDs dataset from machine learning repositories of UCI The dataset used in the study was obtained from Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University and includes 188 patients with PDs (107-males, 81-females) and 64 non-diseased individuals (23-Males and 41-Females). The patient's age ranges were The ages of healthy individuals ranged between 41 to 82 years. The voices were recorded in 44.1 KHz and following doctor's inspections, three duplicates of the vowel /a/ letter of persons were obtained.

3.2. Feature extraction

Features including Baseline features, Time frequency Features, MFCCs (Mel-Frequency Cepstral Coefficients), WTs (Wavelet Transforms), TQWTs (Tunable Q-factor WTs), and Vocal fold features were retrieved from the dataset.

i) Baseline characteristics: Patients affected with PDs experience impaired speech even during early stages of the disease's onset and hence characteristics of speech can be effectively utilised to evaluate PDs and track its progression for medical therapies. Examples of baseline features can be: Jitter and glow-based features, fundamental frequency parameters (#5), harmonicity parameters (#2), RPDEs (Recurrence Time Density Entropies) (#1), DFAs (Detrended Fluctuation Analyses) (#1), and PPEs (Pitch Period Entropies) (#1) are commonly used speech characteristics in Parkinson's disease studies

ii) Time frequency Features: The features like Intensity Parameters (#3), Formant Frequencies (#4) and Bandwidth (#4).

iii) MFCCs: MFCCs based extraction methods uses triangular overlapping filter banks to combine cepstral analysis with spectral domain partitions. The data contained 84 MFCCs related parameters generated using mean and standard deviations of the first 13 MFCCs, as well as the signal's log energy's 1st and 2nd [40], apart from the vocal folds, to identify PD effects in the vocal tract (#84).

iv) WTs: WTs are common approaches for making decisions about signals in general, especially ones with minor geographical variances. This approach produces 182 WT-based characteristics for both the approximation and detailed coefficients, including energy, Shannon's and log energy entropy, and Teager-Kaiser energy.

v) TQWTs): TQWTs use three adjustable parameters (Q (Q-factors), r (redundancies), and J (count of levels)) to improve the quality of signal conversion based on signal behaviour Using this dataset, many tests yielded 432 TQWT-related variables

vi) Vocal fold characteristics: Vibrations of voices in Features were examined for noises. The data was used to extract features like GQs (Glottis Quotients) (#3), GNEs (Glottal to Noise Excitations) (#6), VEFRs (Vocal Fold Excitation Ratios) (#7), and EMDs (Empirical Mode Decompositions) features .

vii) Concat characteristics: Concat features combine baseline, vocal fold and time frequency features.

3.3. Reducing Dimensionalities with KPCAs

KPCAs reduce dimensionalities i.e. they reduce linear dimensions of the high dimensional sound waves for examining variances in the case of PDs . Assume that PDs dataset is labelled i,i=1,...N, with each an I being a D-dimensional sound recording characteristics vector. The sound recordings' dimensionality reduction feature vector is found

using this vector. OBEFSs are utilised for feature selection using dimensionality reduced feature vectors.

3.4. Feature selection by OBEFSs

In this work, OBEFSs techniques by aggregating the feature rankings provided by the single feature selectors such as FMBOAs , LFCSAs , and AFAs into a final consensus ranking via correlation.

3.4.1. FMBOAs

FMBOAs are methods for choosing feature sets. FMBOAs select characteristics for samples based on the degree of existence effects. MBOs are based on the migratory behaviors. FMBOAs have been exploited for good results in classifications when employed without modifications. This implies that they manage to balance their searches both global and local. Butterflies relate and distribute information locally within the swarm and thus assisting in enhanced information for systems where the search operations and sharing of information are algorithmically executed using operators for migrations and butterfly adjustments.

3.4.2. LFCSAs

CSAs (Cuckoo Search Algorithms) are based on cuckoo species which use suitable hosts to optimise selection of attributes from datasets in order to nurture their brood. Some cuckoo species employ a suitable host to optimise the selection of features from a dataset in order to nurse their brood. CSAs are used to reduce the risk of egg losses (irrelevant features) to other species without avoiding parental commitments in raising their progeny. The final qualities are determined by putting eggs (features) in several nests. LFCSAs start with a population of N host nests .A cuckoo's egg (feature), say i, is picked at random at each iteration t, and new feature solutions f_i^{t+1} are generated. The Lévy flights are a sort of random walks where steps are characterized in terms of step-lengths and following certain probability distributions. The step's orientations must be isotropic and random. Probabilities prb_a are used to compute replacement rates in a stochastic manner which are tuned for better performances. The best solutions (features) achieved so far are saved as feature vectors f_{best}, and all current solutions (feature selection solutions) are ordered according to their fitnesses (accuracies) at iterations. The procedure is repeated until the specified stopping threshold is reached.

3.4.3. AFAs

The idealised behaviour of firefly flashing is the basis for FAs Fireflies steps need to placed far away from optimal answers as possible. For the optimal selection of attributes from datasets, fireflies are employed and global and local searches are balanced. As a result, firefly's strides also need to consider past and current location's data. Historical data of Fireflies which contain optimal values of past two iterations are taken into account in this work. The distances between current fitness values and population's best fitness values determine further steps of fireflies in iterations. The steps may vary with iterations, and steps of fireflies are also altered in iterations.

3.4.4. Correlation function

Correlation coefficient matrices from features are computed for selecting outputs of ensemble's feature selection and use Equation (1),` vv`

$$Correlation\ Coefficient = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}} \quad (1)$$

Where *x, y* represent examined attribute values while *N* stands for instance counts. Feature sets chosen based on correlations by ensembles become inputs for classifiers.

3.5. classifications of PDs via Ensemble Deep Learning (EDL) classifier

In this work, PDs are classified using an Ensemble Deep Learning (EDL) classifier. Ensemble learning can considerably boost the learning system's generalisation capabilities. EDL, a machine-learning approach that uses numerous classifiers such as FCBi-LSTMs, CAEs, and SAEs to construct an ensemble learner using Stacked Generalization to enable greater generalisation of learning systems, has shown remarkable success in classifications of PDs . Deep EDL combines the benefits of three deep learning models as well as ensemble learning to provide a model with improved generalisation performance. The EDL classifier learns numerous base classifiers and aggregates their results using specified criteria. The rule used to aggregate the outputs defines an ensemble's effective performance..

3.5.1. FCBi-LSTMs

The FCBi-LSTMs are used for classifications based on characteristics chosen from PD datasets . Convolution and pooling layers make up these networks. They carry out convolutions then pooling where outputs are provided as inputs for subsequent Convolution layers. Convolution layers using filters divide features into multiple matrices of sizes (Nm+1) for better representations. These resultant matrices are reduced in dimensions by CNN's pooling layers but manage to maintain links between features. These averaging layers which use preceding Convolved inputs, send their outputs as inputs to subsequent layers. Bi-LSTMs use final Convolution layer's outputs as intermediary variables and hence perform better than uni-directional LSTMs and retain more structural information. The final outputs of Bi-LSTMs are processed by two Convolution layers to obtain diagnosis of PDs. MFB (Multi-modal Factorized Bilinear Pooling) is a technique for combining information from CNNs with Bi-LSTM's obtained features.

3.5.2. SAEs

The SAEs classifier neurons labelled as (+1) are bias units introduced to the FFNNs (feed-forward NNs) through cost functions. This phase drives AEs more correctly to replicate inputs x without over fits .Figure 2 depicts an overview of the model. The sparse auto encoder bottlenecks are utilised as input vectors to DNNs .

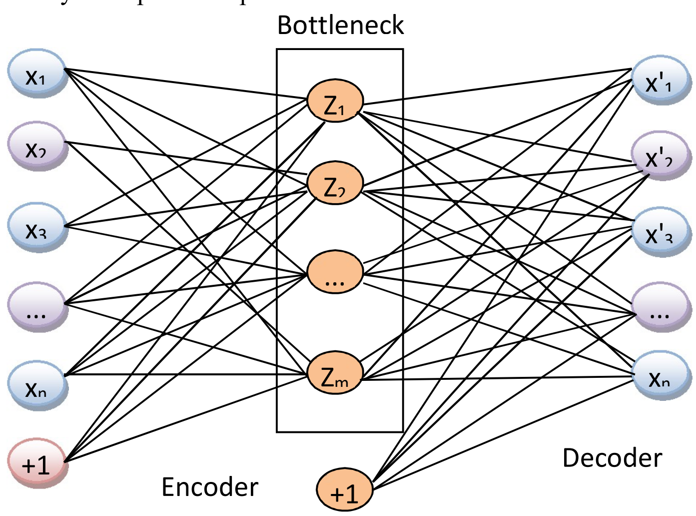


Figure 2. Sparse Model Of Saes

SAEs are proposed in which cost functions are made of 3 parts and are detailed below. Assuming dataset has N training instances $(x_1, x_2, \dots x_n)$, where x_i stands for ith input. SAEs learn to reconstruct inputs x_i using cost functions $h_{W,b}(x_i)$ nearing x_i where MSEs, weight decays, and sparsity are included in the function. The cost functions for N training samples MSEs and weight decays are defined in the following equation [53],

$$J_{sparse}(W,b) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left\|h_{W,b}(x^i) - x^i\right\| + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^l\right)^2 \quad (2)$$

Weight decays specified by the equation above avoids over fits of data, since small values of λ can lead to data over fits, while its larger values can result in data under fits. Sparsity, the third component of the cost functions, helps in activating hidden layers of AEs and prevent data over fits. It has the ability to minimise the number of hidden layer areas evaluated. The average active value of the hidden layer was calculated using the following equation, where a denotes the activation function, which is rectifier (ReLU),

$$\hat{p}_j = \frac{1}{N}\sum_{i=1}^{N}\left(a_j^2(x^i)\right) \quad (3)$$

Sparsity is computed for getting $\hat{p}_j$ nearer to p (sparsity parameter) helps in deviations from p and results in activating or de-activating hidden layer's neurons. It can be defined using Kullback-Leibler divergences as depicted in Equation (4)

$$\sum_{j=1}^{s_l} KL(p\|\hat{p}_j) = \sum_{j=1}^{s_l}\left[p\log\frac{p}{\hat{p}_j} + (1-p)\log\frac{1-p}{1-\hat{p}_j}\right] \quad (4)$$

SAEs cost functions add all three component results which is depicted as Equation (5),

$$J_{sparse}(W,b) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left\|h_{W,b}(x^i) - x^i\right\| + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^l\right)^2 + \beta\sum_{j=1}^{s_l}KL\left(p\|\hat{p}_j\right) \quad (5)$$

where β represents sparse penalties. The deep NNs classifier is trained using the bottleneck of the SAEs as inputs, and the SAEs are trained to minimise their cost function as mentioned above. The SAEs and the classifier were both trained at the same time, resulting in improved feature extraction while the classifier's choice was optimised. The training procedure takes 30 iterations, using an 8 instances per batch. The sparsity parameter p was set to 0.05 while weight decay λ was set to 0.0001and sparse penalty term to 2. DNN fine-tuning on the final 10 iterations to change classifier parameters and reduce the softmax cost function while SAE parameters remain fixed. The parameters are updated using the Adam optimizer depending on the calculated gradients.

3.5.3. CAEs

Contractive AEs transform learned representations are robust towards small changes around the training examples It achieves that by using different penalty term imposed to the representation. The loss function (ℓ_2) is used for the reconstruction term. AEs use the encoder function f to map inputs x to internal representations (or codes) f(x). The decoder functions g are then used to map f(x) back to their input spaces. Reconstruction functions r made up of the functions f and g i.e. r(x) = g(f(x)), and reconstruction loss functions l penalises errors, with r(x) considered as forecasts of x. CAEs [54] are regularised AEs that learn to minimise the regularised reconstruction errors and depicted below.

$$\mathcal{L}_{CAE} = \mathbb{E}\left[\ell(\mathrm{x}, \mathrm{r}(\mathrm{x})) + \lambda\left\|\frac{\partial \mathrm{f}(\mathrm{x})}{\partial \mathrm{x}}\right\|_{\mathrm{F}}^2\right] \quad (6)$$

where $r(x) = g(f(x))$ and $\|A\|_F^2$ is the sum of the squares of the elements of A. Both the squared loss,

$$\ell(\mathrm{x}, \mathrm{r}) = \|\mathrm{x} - \mathrm{r}\|^2 \quad (7)$$

and the cross-entropy loss,

$$\ell(\mathrm{x}, \mathrm{r}) = -\mathrm{x}\log \mathrm{r} - (1 - \mathrm{x})\log(1 - \mathrm{r}) \quad (8)$$

Because of the simpler mathematical method it enables, concentrate your investigation on the squared loss. It's worth noting that minimising the CAEs criteria is very dependent on the parameterization of f and g, particularly the linked weights restriction imposed by the equation below (9),

$$\mathrm{f}(\mathrm{x}) = \mathrm{sigmoid}(\mathrm{Wx} + \mathrm{b}) \quad (9)$$

$$g(h) = sigmoid(W^T h + c) \quad (10)$$

Because of the linked weights, the foregoing regularising term compels f (as well as g) to be contractive, i.e., to have singular values < 1. Larger values of λ produce greater contractions (smaller singular values) where it has least impacts on reconstruction errors, i.e. in local directions with little or no variability of data.

3.5.4. Stacked Generalization

Stacked generalisation deduces generalized biases in relation to given learning sets. Cross validation data and least squares with non-negativity constraints were used to determine the best weights of combination in regression to generate a good linear combination of the base learners . Consider the linear combination of the base learners' predictions $f_1, f_2, \ldots, f_m$ given by equation (11),

$$f_{stacking}(x) = \sum_{j=1}^{m} w_j f_j(x) \quad (11)$$

where w is the optimal weight vector learned by the meta learner

4. EXPERIMENTAL RESULTS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This work's experimental findings of the suggested EDL classifier and its comparisons with other approaches including FCLSTM-CNNs (Fuzzy Convolution Long Short-Term Memories) based CNNs, FCBi-LSTMs, and CNNs are detailed in this section. MATLAB R2016a was used to conduct the studies on recognising PDs and classifying them. The following system requirements were followed during implementation: Intel(R) Core™i3-4160T CPU@3.10 GHz 3.09 GHz processor, 4.00 GB RAM, Windows 8.1 pro, 64 bit operating system, operation system, and 1 TB hard disk.

4.1. EVALUATION METRICS

To analyse prediction performances of classifiers, evaluation metrics are required. Confusion matrix in Table 1 represents the counts of properly and wrongly categorised occurrences for classes based on binary classifications. The symbols tp, fp, fn, and tn in the confusion matrix represent true positive (tp), false positive (fp), false negative (fn), and true negative (tn), respectively. Precisions, recalls, F-measures, accuracies, and error rates were computed using formulae given in Equations (12-16),

$$Precision = \frac{tp}{tp + fp} \qquad (12)$$

$$recall = \frac{tp}{tp + fn} \qquad (13)$$

$$F - measure = \frac{2 * precision * recall}{precision + recall} \qquad (14)$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \qquad (15)$$

Error Rate=100-Accuracy (16)

Another statistic for evaluating the quality of binary classifications is the MCCs . MCCs consider tp, fp, fn, and tn counts and are recognised as balanced measures that can be applied even when class distributions are imbalanced. MCCs are correlation coefficients between actual and projected occurrences that range in the interval [-1, +1] where +1 represents flawless predictions and -1 shows differences between forecasts and actual in labels.

Table 1. Confusion Matrix For Two-Class Classification

ACTUAL/ PREDICTED AS	POSITIVE	NEGATIVE
POSITIVE	tp	fn
NEGATIVE	fp	tn

4.2. RESULTS COMPARISON

Experimentations using 4 features and classifiers were evaluated using the metrics of accuracies, errors, F-measures, and MCCs. The combination of MFCC + Wavelet + Concat features with a CNNs classifier yields an accuracy rating of 94.1752 percent. Despite the fact that the accuracy percentage of the MFCC + Wavelet + Concat combination is 95.1557 percent for the FCLSTM-C FCBi-LSTM classifier gives the accuracy results of 96.6381%, 98.0244%, 97.3457% and 98.7720% respectively for TQWT+MFCC+Wavelet, TQWT+Wavelet+Concat, TQWT+MFCC+Concat, and MFCC+Wavelet+Concat. Proposed EDL classifier with MFCC + Wavelet + Concat combination achieves the highest performance with an accuracy rate of 99.903%, F-Measure rate of 99.633%, and 72.431% for MCCs (See Table 2).

Table 2. Results Of Classifiers With Triple Feature (Kpca+ Obefs)

CNNs CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR RATE**	**MCCs**
F1-TQWT+MFCC+Wavelet	85.8697	87.5696	12.4304	57.2007
F2-TQWT+MFCC+Concat	90.2315	91.9315	8.0684	61.4600
F3-TQWT+ Wavelet + Concat	86.3695	88.0694	11.9306	63.3994
F4-MFCC + Wavelet + Concat	92.4752	94.1752	5.8248	64.5384
FCLSTM-CNNs CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR RATE**	**MCCs**

F1-TQWT+MFCC+Wavelet	94.2258	93.0470	6.9530	67.6669
F2-TQWT+MFCC+Concat	91.5250	93.0854	6.9146	67.7060
F3-TQWT+ Wavelet + Concat	93.4200	93.1261	6.8739	65.4457
F4-MFCC + Wavelet + Concat	91.6921	95.1557	4.8443	67.2960
FCBi-LSTM CLASSIFIER(%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR RATE**	**MCCs**
F1-TQWT+MFCC+Wavelet	98.3100	96.6381	3.3619	74.300
F2-TQWT+MFCC+Concat	96.5900	98.0244	1.9756	72.300
F3-TQWT+ Wavelet + Concat	97.5200	97.3457	2.6543	70.300
F4-MFCC + Wavelet + Concat	98.5010	98.7720	1.2280	71.400
EDL CLASSIFIER (%)				
FEATURE COMBINATION	**F-MEASURE**	**ACCURACY**	**ERROR RATE**	**MCCs**
F1-TQWT+MFCC+Wavelet	99.449	97.771	2.2299	75.432
F2-TQWT+MFCC+Concat	97.722	99.156	0.8436	73.412
F3-TQWT+ Wavelet + Concat	98.652	98.378	1.6222	71.013
F4-MFCC + Wavelet + Concat	99.633	99.903	0.0966	72.431

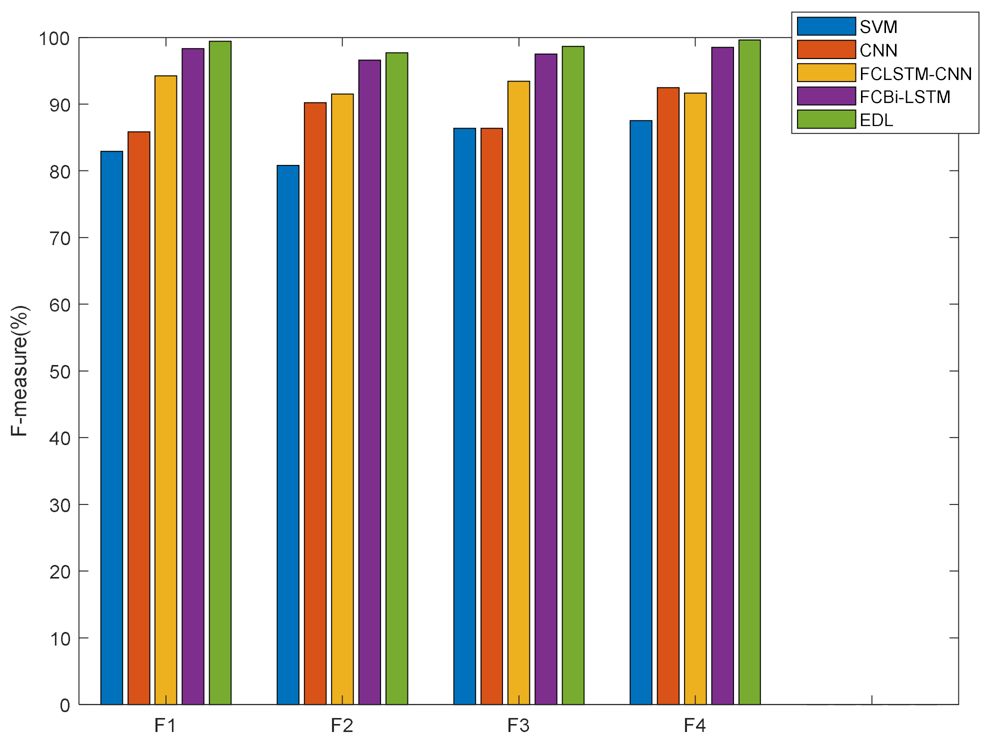


Figure 3. F-Measure Comparison Of Feature Level Combination Vs. Classifiers

Figures 3-6 shows F-measures, accuracies, MCCs , and errors using different feature sets. When compared to other feature combination sets, TQWT+MFCC+Wavelet combination of feature set yields greater results of 99.449 percent, 97.77 percent, 75.432 percent, and 2.2299 percent for f-measure, accuracy, MCCs , and error. The suggested classifier using the first feature level combination achieves a higher f-measure of 99.449 percent, whereas other approaches such as CNNs , FCLSTM-CNNs , and FCBi-LSTM achieve f-measures of 85.8697 percent, 94.2258 percent, and 98.3100 percent, respectively.

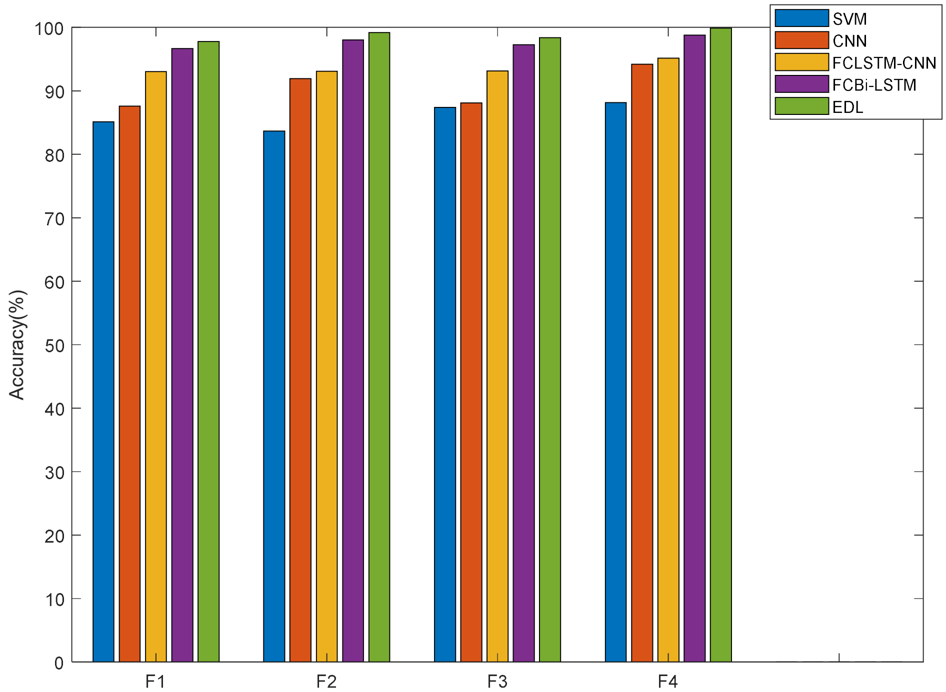


Figure 4. Accuracy Comparison Of Feature Level Combination Vs. Classifiers

Figure 4 shows that the x-axis results are measured using a feature-level combination of classifiers, and the y-axis results are shown as accuracy. The final feature level combination proposed for the classifier achieves 99.9030 percent accuracy, whereas other approaches such as CNNs , FCLSTM-CNNs , and FCBi-LSTM achieve 94.1752 percent, 95.1557 percent, and 98.7720 percent accuracy, respectively.

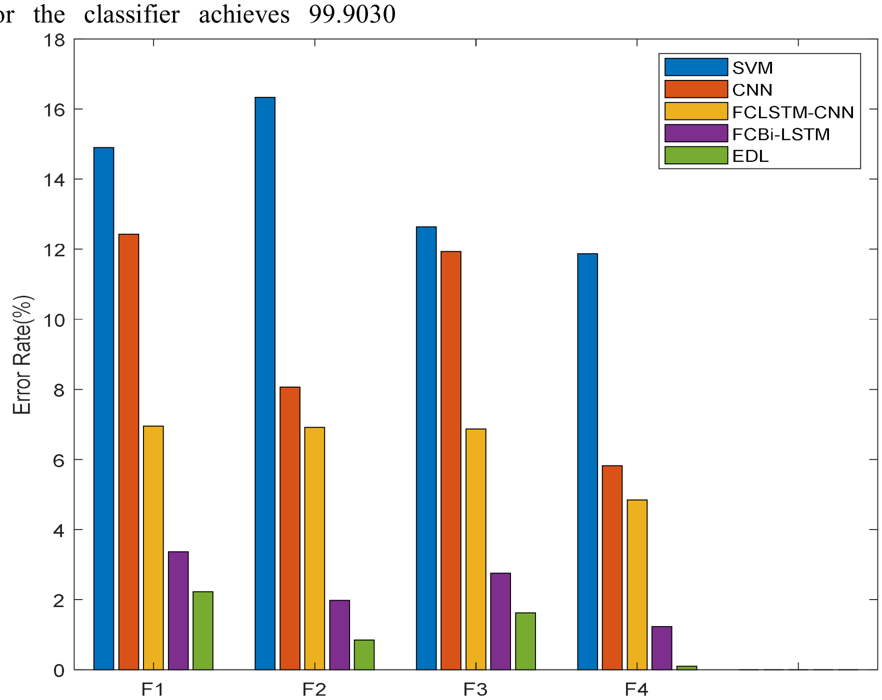


Figure 5. Error Rate Comparison Of Feature Level Combination Vs. Classifiers

Figure 5 compares the error rates of classifiers with four different feature level combinations. Figure 5 shows that the proposed classifier with the final feature level combination has a lower error rate of 0.0966 percent, whereas CNNs , FCLSTM-CNNs , and FCBi-LSTM approaches have larger error rates of 5.8248 percent, 4.8443 percent, and 1.2280 percent, respectively.

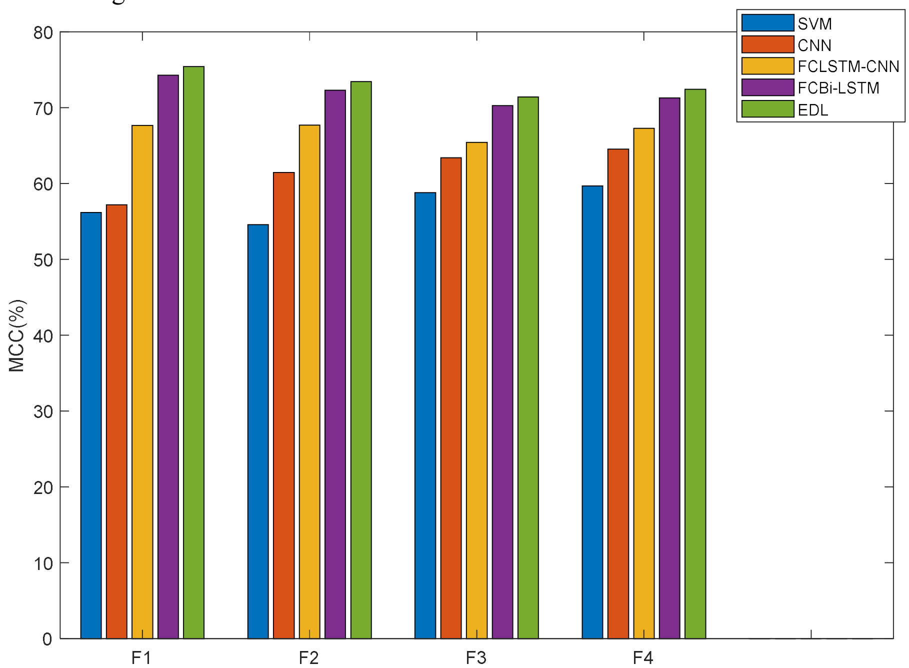


Figure 6. Mccs Comparison Of Feature Level Combination Vs. Classifiers

Figure 6 displays MCCs results comparison of first feature level combination produces greater results of 75.432 percent for proposed algorithm, 57.2007 percent , 67.6669 percent , and 74.300 percent for CNNs , FCLSTM-CNNs , and FCBi-LSTM approaches classifiers accordingly.

5. CONCLUSION AND FUTURE WORK

In this study, an ensemble learner known as an OBEFSs is proposed along with Ensemble Deep Learning (EDL) classifier that combines single approaches. This is done using algorithms such as FMBOAs, LFCSAs, and AFAs. To carry out OBEFS, the correlation function is used to choose the best features from the three feature subsets. After that, the EDL classifier is utilised to get a diagnostic of patients with PDs. The EDL classifier contained FCBi-LSTMs , CAEs , and SAEs . The stacked generalisation method was developed to aggregate the findings of classifiers that only choose the majority class (PDs) of the dataset. F-Measures, Accuracies, MCCs , and error rates were used to evaluate the categorization The classifier outputs are implemented using machine learning repositories of UCI. Experimentation is carried out with four different types of feature sets and classifiers. The proposed EDL classifier using MFCC+Wavelet+Concat combination obtains the greatest accuracy rate of 99.9030 percent, F-Measure rate of 99.633 percent, and MCCs rate of 72.431 percent. Most PDs research rely only on accuracy rates, which might be deceptive in cases of skewed class distribution. Instead, various statistical techniques like as cross-validation and the ANOVA test have been included to examine the classifier outcomes.

DECLARATION:

Ethics Approval and Consent to Participate:

No participation of humans takes place in this implementation process

Human and Animal Rights:

No violation of Human and Animal Rights is involved.

Funding:

No funding is involved in this work.

Data availability statement:

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study

Conflict of Interest:

Conflict of Interest is not applicable in this work.

Authorship contributions:

All authors are contributed equally to this work

Acknowledgement:

There is no acknowledgement involved in this work.

REFERENCES

[1]. H.Gao, X. Wei, J. Liao, R. Wang and J. Xu, "Lower bone mineral density in patients with Parkinson's disease: a cross-sectional study from Chinese Mainland," *Frontiers in aging neuroscience*, vol.7,2015, pp.1-9. https://doi.org/10.3389/fnagi.2015.00203 .

[2]. S.Halbgebauer, M. Nagl, H. Klafki, U.Haußmann, P. Steinacker, " Modified serpinA1 as risk marker for Parkinson's disease dementia: Analysis of baseline data,"*Scientific reports*, vol.6,2016, pp.1-8. DOI: 10.1038/srep26145

[3]. X.Wei, H. Gao, J. Zou, X. Liu, D. Chen , » Contra-directional coupling of Nur77 and Nurr1 in neurodegeneration: a novel mechanism for memantine-induced anti-inflammation and anti-mitochondrial impairment,"*Molecular neurobiology*, vol.53,2016, pp.5876-5892. DOI: 10.1007/s12035-015-9477-7

[4]. P.Ashok Babu, L. Kavisankar, J. Xavier, V. Senthilkumar, G. Kumar, T. Kavitha, A. Rajendran, G. Harikrishnan, A. Rajaram and A.G.Adigo, "Selfish Node Detection for Effective Data Transmission Using Modified Incentive Sorted Pathway Selection in Wireless Sensor Networks,"*Wireless Communications and Mobile Computing*, *2022*. https://doi.org/10.1155/2022/9359135.

[5]. J.F.Guo, K. Li, R.L. Yu, Q.Y. Sun, L. Wang, L.Y. Yao, " Polygenic determinants of Parkinson's disease in a Chinese population," *Neurobiology of aging*, vol.36,2015, pp.1765-e1. DOI:10.1016/j.neurobiolaging.2014.12.030

[6]. M.S.Bryant, D.H. Rintala, J.G. Hou and E.J.Protas, " Relationship of falls and fear of falling to activity limitations and physical inactivity in Parkinson's disease,"*Journal of aging and physical activity*,vol. 23,2015, pp.187-193. DOI:10.1123/japa.23.2.187

[7]. A.Rajaram and K.Vinothkumar, " Efficient Multipath Location Aware Routing Protocol for Mobile Ad Hoc Networks," *Journal of Theoretical and Applied Information Technology, Scopus Indexed*, Vol.*59,2019,* pp.130-38.

[8]. M.A.Nalls, N. Pankratz, C.M. Lill, C.B. Do, " Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease,"*Nature genetics*, vol.46,2014, pp.989-993. DOI:10.1038/ng.3043

[9]. H.Gürüler H., "A novel diagnosis system for parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method," *Neural Computing and Applications*, vol. 28, 2017, pp. 1657–1666. DOI:10.1007/s00521-015-2142-2

[10]. A.Rajaram, and A.Baskar, "Hybrid Optimization-Based Multi-Path Routing for Dynamic Cluster-Based MANET," *Cybernetics and Systems*.2023. https://doi.org/10.1080/01969722.2023.2166249

[11]. A.Rajaram and J.Sugesh, "Power aware routing for MANET using on-demand multipath routing protocol," *International Journal of Computer Science Issues (IJCSI)*, vol.*8, 2011,pp*.517. DOI:10.1109/ICDCSW.2004.1284112

[12]. M.Peker, " A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM," *Journal of medical systems*, 40,2016, pp.1-16. DOI:10.1007/s10916-016-0477-6

[13]. B.E.Sakar G. Serbes, and C. O. Sakar, "Analyzing the effectiveness of vocal features in early telediagnosis of parkinson's disease," *PloS one*, vol. 12, 2017, pp.1-18. DOI:10.1371/journal.pone.0182428.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[14]. K.J.Kubota , J. A. Chen and M. A. Little, "Machine learning for largescale wearable sensor data in parkinson's disease: Concepts, promises, pitfalls, and futures," *Movement disorders*, vol. 31, 2016, pp. 1314–1326. DOI:10.1002/mds.26693.

[15]. D.N.V.S.L.S. Indira, R.K., Ganiya, P. Ashok Babu, A. Xavier, L. Kavisankar, S. Hemalatha, V. Senthilkumar, T. Kavitha, A. Rajaram, K. Annam and A.Yeshitla, " Improved artificial neural network with state order dataset estimation for brain cancer cell diagnosis." *BioMed Research International*, *2022*. https://doi.org/10.1155/2022/7799812

Avinashilingam Institute for Home Science and Higher Education for Women
(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category A by MHRD)
Re-accredited with 'A++' Grade by NAAC.CGPA 3.65/4, Category I by UGC
Coimbatore - 641 043, Tamil Nadu, India

PLAGIARISM CHECK REPORT (THESES)

1.	Name of the Research Scholar	B. Sabeena
2.	Roll No. and Year of Registration	19PHEOF002, 2019
3.	Department	Computer Science and Engineering
4.	Name of the Research Guide	Dr. S. Sivakumari
5.	Title of the Thesis / Dissertation	Study on the Performance of Deep Learning Techniques for the Classification of Parkinson's Diseases
6.	Similarity Content (%) Identified	9%
7.	Software Used	Turnitin
8.	Date of Verification	26-02-2024

Note : The report is excluding 14 Consecutive words, Review of Literature and Quoted Materials.

Checked by :

for K. [signature]
Information Scientist

[signature]
Research Scholar

[signature] 26.02.24
Assistant Librarian

[signature] 26/2/24
Research Guide

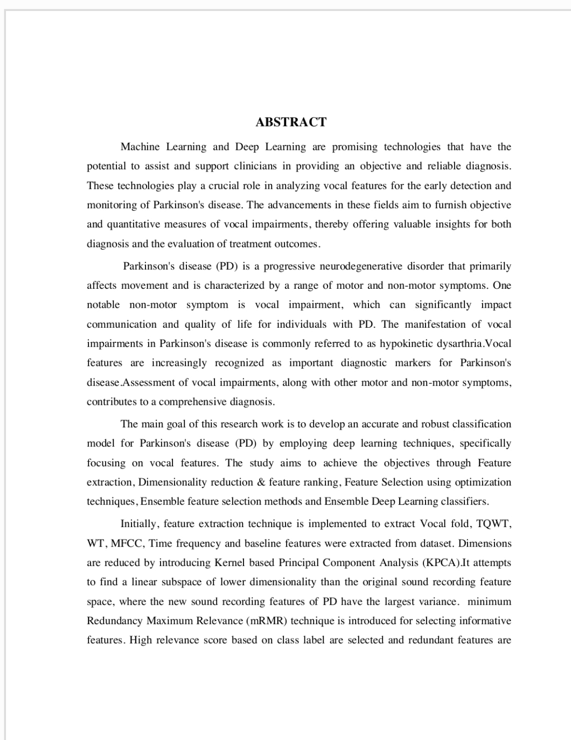
Date: 26-02-2024

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Library Adu
Assignment title:	New 2022
Submission title:	Study on the Performance of Deep Learning Techniques for ...
File name:	sabeena_thesis_25.2.2024_plagirism_check_document.docx
File size:	3.03M
Page count:	121
Word count:	24,951
Character count:	145,734
Submission date:	26-Feb-2024 02:07PM (UTC+0530)
Submission ID:	2304829210



ABSTRACT

Machine Learning and Deep Learning are promising technologies that have the potential to assist and support clinicians in providing an objective and reliable diagnosis. These technologies play a crucial role in analyzing vocal features for the early detection and monitoring of Parkinson's disease. The advancements in these fields aim to furnish objective and quantitative measures of vocal impairments, thereby offering valuable insights for both diagnosis and the evaluation of treatment outcomes.

Parkinson's disease (PD) is a progressive neurodegenerative disorder that primarily affects movement and is characterized by a range of motor and non-motor symptoms. One notable non-motor symptom is vocal impairment, which can significantly impact communication and quality of life for individuals with PD. The manifestation of vocal impairments in Parkinson's disease is commonly referred to as hypokinetic dysarthria.Vocal features are increasingly recognized as important diagnostic markers for Parkinson's disease.Assessment of vocal impairments, along with other motor and non-motor symptoms, contributes to a comprehensive diagnosis.

The main goal of this research work is to develop an accurate and robust classification model for Parkinson's disease (PD) by employing deep learning techniques, specifically focusing on vocal features. The study aims to achieve the objectives through Feature extraction, Dimensionality reduction & feature ranking, Feature Selection using optimization techniques, Ensemble feature selection methods and Ensemble Deep Learning classifiers.

Initially, feature extraction technique is implemented to extract Vocal fold, TQWT, WT, MFCC, Time frequency and baseline features were extracted from dataset. Dimensions are reduced by introducing Kernel based Principal Component Analysis (KPCA).It attempts to find a linear subspace of lower dimensionality than the original sound recording feature space, where the new sound recording features of PD have the largest variance. minimum Redundancy Maximum Relevance (mRMR) technique is introduced for selecting informative features. High relevance score based on class label are selected and redundant features are

Study on the Performance of Deep Learning Techniques for the Classification of Parkinson's Diseases

by Library Adu

Submission date: 26-Feb-2024 02:07PM (UTC+0530)
Submission ID: 2304829210
File name: sabeena_thesis_25.2.2024_plagirism_check_document.docx (3.03M)
Word count: 24951
Character count: 145734

Study on the Performance of Deep Learning Techniques for the Classification of Parkinson's Diseases

ORIGINALITY REPORT

9%	7%	7%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.jatit.org Internet Source	1%
2	www.ncbi.nlm.nih.gov Internet Source	1%
3	tojqi.net Internet Source	1%
4	9lib.net Internet Source	1%
5	Shriniket Dixit, Khitij Bohre, Yashbir Singh, Yassine Himeur, Wathiq Mansoor, Shadi Atalla, Kathiravan Srinivasan. "A Comprehensive Review on AI-Enabled Models for Parkinson's Disease Diagnosis", Electronics, 2023 Publication	<1%
6	downloads.hindawi.com Internet Source	<1%
7	www.researchgate.net Internet Source	<1%