

ABSTRACT

Coronavirus disease 2019 (COVID-19) is a new global health emergency with several infected cases and deaths recorded all over countries. On the other side Pneumonia, which affect respiratory organs and affects its function. A premature forecast of COVID-19 and Pneumonia are significant for patient isolation and rapid clinical involvement. Only some techniques have been designed for the forecast of COVID-19 and Pneumonia. However, most of the existing methods have failed to provide accurate disease prediction within a minimum time frame. With this intent, novel techniques are introduced in the proposed research work to achieve higher accuracy and reduce the time required for preprocessing, feature selection, and classification.

In the initial phase of research work, two novel preprocessing techniques are introduced for preprocessing the COVID-19 coronavirus and RSNA Pneumonia datasets, achieving superior accuracy and reduced time. Initially, the Additive Log Ratio Transformed One Hot Encoding (ALRTOHE) Technique is proposed to improve the accuracy of preprocessing with minimum complexity. After that, the Zero Mean Feature Normalized Encoding (ZMFNE) technique is designed to preprocess the coronavirus data with maximum accuracy. Zero Mean Feature scaling is applied to normalize input information. Then, the data transformation process is accomplished by encoding the data into a binary representation using one-hot encoding to reduce complexity.

After preprocessing, three novel feature selection techniques are developed in the second research work to enhance disease forecasting results through a lower error rate. First, the Non-Linear Sammon Projective Pattern Selection (NSPPS) Model is designed to select pertinent features for disease forecasting. Secondly, the Tversky Similarity Indexed Distributive Feature Embedding Technique (TSIDFE) method is developed to select the most relevant features with maximum accuracy for predicting the disease. Third, the Statistical Correlative Targeted Projection Pursuit-Based Feature Selection (SCTPP-FS) method is proposed to extract pertinent features from the dataset for improved pulmonary disease diagnosis.

To achieve efficient classification, Machine Learning (ML) and Deep Learning (DL)- based methods are introduced in the third research work for accurate COVID-19 and Pneumonia disease forecasting. In the primary process, the Emphasis Perceptron

Boosting Classification (EPBC) technique is developed to perform robust disease prediction with improved accuracy and reduced time. The time-dependent Cox regressive Levenberg-Marquardt Convolutional Neural Learning (TCLMCNL) Technique is designed to categorize patient data for robust disease prediction. A Memetic Optimized U-Net Deep Learning (MO-UNetDL) classifier method is proposed for precise disease forecasting through improved precision and recall.

The proposed preprocessing techniques, such as ALRTOHE and ZMFNE, feature selection techniques, including NSPPS, TSIDFE, and SCTPP-FS, and classification techniques, including EPBC, TCLMCNL, and MO-UNetDL, are implemented using the Python programming language. The experiments are conducted using the COVID-19 Corona Virus India Dataset and the RSNA Pneumonia Detection Challenge dataset, with varying numbers of patient information samples as input. Experimental results of preprocessing techniques ALRTOHE and ZMFNE are measured in terms of preprocessing accuracy, preprocessing time, space complexity, and error rate, compared to existing methods such as the Convolutional Neural Network - Gated Recurrent Units based hybrid Deep Learning (CNN-GRU) method and a Variant of Concerned Deep Learning (VOC-DL). In comparison to other methods, the proposed ZMFNE achieved better performance using the COVID-19 Corona Virus India Dataset, increasing preprocessing accuracy by 13%, preprocessing time by 17%, space complexity by 18%, and reduced error rate by 48% compared to conventional techniques. The ZMFNE method of performance preprocessing is improved using the RSNA Pneumonia Detection Challenge dataset compared to the ALRTOHE method, achieving an accuracy of 89.8%, a 16.23% of error rate, a 65KB space complexity, and a 5090 ms of preprocessing time.

The performance of feature selection techniques, such as NSPPS, TSIDFE, and SCTPP-FS, is compared with that of the existing AHEG-FS method and a hybrid Chi2-MI-based feature selection model. In comparison to other existing feature selection techniques, the results of SCTPP-FS demonstrate improved feature selection accuracy by 11%, reduced feature selection time by 26%, decreased space complexity by 24%, and a 48% reduction in error rate, surpassing existing techniques. The results of the SCTPP-FS method are presented using the RSNA Pneumonia Detection Challenge dataset, which enhances accuracy to 94%, reduces the error rate to 6%, achieves a space complexity of 50KB, and decreases the feature selection time to 3070ms.

Similarly, the output of proposed classification methods, such as EPBC, TCLMCNL, and MO-UNetDL, is compared with existing Deep learning LSTM ensemble and DSPM models. The MO-UNetDL technique outperformed others in the COVID-19 Corona Virus India database in terms of prediction accuracy by 3%, precision by 4%, recall by 3%, specificity by 21%, and prediction time by 27% compared to existing methods. The outcome of the MO-UNetDL technique yields better classification results using the RSNA Pneumonia Detection Challenge dataset, with an accuracy of 98%, precision of 98.5%, recall of 97.65%, and specificity of 59.25%, as well as a prediction time of 2056ms.