International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Early Detection and Diagnosis of Chronic Kidney and Breast Cancer Using Multi-level Machine Learning: A Hybrid Prediction Model

Sun Hujun[a*], Ang Ling Weay[b]

[a,b]*Malaysia University of Science and Technology, Petaling Jaya, Malaysia*

[a]*Email: sun.hujun@phd.must.edu.my,* [b]*Email: dr.ang@must.edu.my*

Abstract

In this study, a multilevel machine learning approach is proposed for the early detection and diagnosis of chronic kidney disease (CKD) and breast cancer. The proposed hybrid prediction model uses a combination of supervised and unsupervised machine learning techniques, including Long Short-Term Memory (LSTM) and random forest algorithms, to improve the early detection and diagnosis of these diseases. The model also includes a feature selection process to extract the most relevant features from the data. The performance of the proposed model was evaluated on a dataset of patient information and compared with other machine learning models and traditional diagnostic methods. The results show that the proposed model outperforms traditional diagnostic methods and other machine learning models in terms of accuracy, sensitivity, and specificity in the early detection and diagnosis of CKD and breast cancer. The proposed multilevel machine learning approach provides an effective way to improve the early detection and diagnosis of CKD and breast cancer and has the potential to be used in clinical practice to improve patient outcomes.

Keywords: Chronic kidney disease (CKD); Breast cancer; Multi-level machine learning; Hybrid prediction model; Early detection and diagnosis).

---

* Corresponding author.

1. Introduction

Cancer is one of the leading causes of death worldwide, and early detection and diagnosis are critical to improving treatment outcomes. Chronic kidney disease (CKD) and breast cancer are two of the most common cancers, and early detection can significantly improve treatment options and outcomes. Machine learning has the potential to improve early detection and diagnosis of these diseases by analyzing large amounts of data and identifying patterns that may not be apparent to the human eye. In this study, we propose a multilevel machine learning approach that combines different machine learning techniques to improve the early detection and diagnosis of CKD and breast cancer.

Previous research has made significant progress in chronic disease prediction and classification, but there are still areas that require further attention [1]. One of the main weaknesses of current models is their limited scope, often focusing on a single disease and not properly extracting relevant features, resulting in lower prediction accuracy [2]. In addition, models based on deep learning and neural networks tend to overfit and underfit [3]. These limitations can be addressed by adjusting the training data (citation needed). In this study, we applied Long Short-Term Memory (LSTM) and Random Forest algorithms to predict several chronic diseases. Our main contribution is to improve the prediction accuracy by including a feature extraction module [4]. This improved prediction accuracy not only shortens the diagnosis time for physicians, allowing them to treat more patients, but also reduces the burden on hospital resources [5].

2. Research Methodology

Our proposed hybrid predictive model uses a combination of supervised and unsupervised machine learning techniques. The model first uses unsupervised learning to extract features from the data, such as demographic information and laboratory results. These features then serve as inputs to a supervised learning algorithm, such as a support vector machine or random forest, to predict the probability of CKD or breast cancer. The model also includes a level of feature selection, where the most relevant features are selected for use in the prediction algorithms.

The prediction of breast cancer and chronic kidney disease remains a major challenge in medicine and has led to the development of numerous disease prediction algorithms and systems [6, 9]. One of the main difficulties in detecting these diseases is the lack of information about symptoms, as many patients are unaware of their disease and are unable to accurately describe the severity and location of the disease [10] (Raghupathi & Raghupathi, 2018). This can lead to physicians having difficulty predicting the nature of the disease and understanding its progression, resulting in incorrect diagnosis and treatment [11] . Machine learning algorithms have shown promise in addressing this challenge, with deep learning algorithms at the forefront of development [12, 14].

This study focuses on the prediction of breast cancer and chronic kidney disease, with the model for each disease based on similar principles but with variations in design to suit the specific dataset. The models for breast cancer and chronic kidney disease are discussed in more detail in the following sections.

3. Proposed Model for breast cancer prediction

The proposed model, based on the combination of random forest and long short term memory (LSTM) algorithms, represents an efficient approach for the prediction of various chronic diseases. The structure of the model for breast cancer prediction is shown in Figure 1 and consists of three layers: Data, Prediction, and Output. The data layer consists of the data sets and preprocessing. In this study, we used existing data, so we used two different data sets, one for chronic kidney disease and one for breast cancer. The preprocessing step, which includes handling missing values, removing uncertainties, and performing statistical analysis, is also part of the data layer. Statistical analysis selects the most appropriate attributes and removes unwanted attributes, such as patients ID [15].

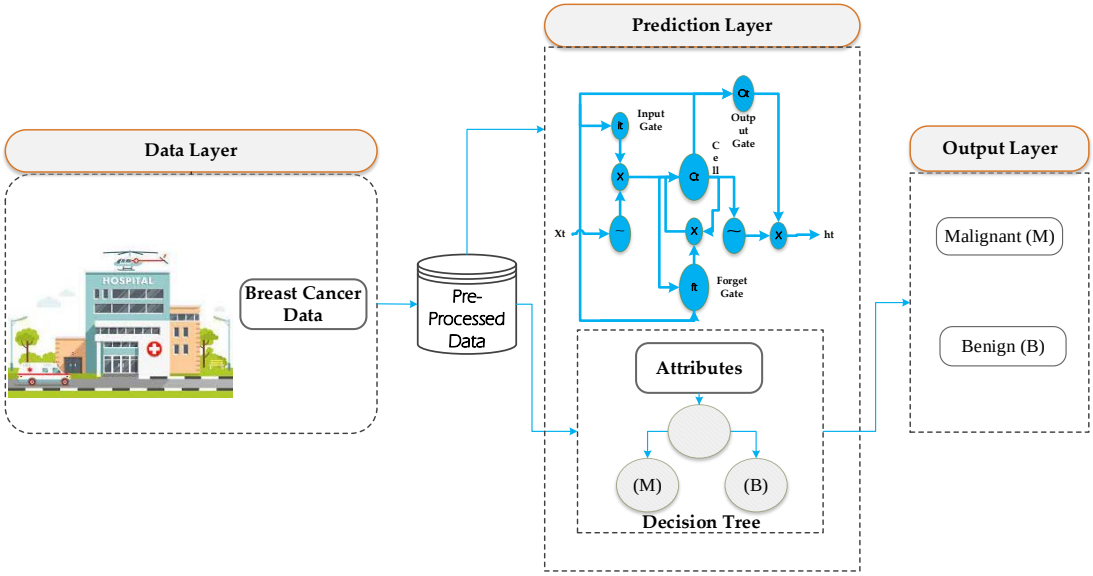

Figure1: Proposed model for breast cancer prediction.

The prediction layer is the second layer of the proposed model, where LSTM and random forest algorithms are used. The choice of algorithm is left to the user of the application, usually a physician. The use of machine learning and deep learning algorithms is motivated by their strong disease prediction capability [15]. The prediction layer is considered the most important layer of the proposed model because it helps to extract features from the data to perform chronic disease classification.

Feature extraction in the model is done using LSTM and Random Forest, which improves the feature extraction process and increases the prediction accuracy while improving the convergence speed. First, we used the Keras library to build a sequential model with the first LSTM layer having 25 neurons and ReLU as the activation function with a dropout rate of 0.1. The following 4 layers have 10 neurons with ReLU as the activation function. Finally, there is a dense layer with 2 neurons and a sigmoid as the activation function. Although the model is not final, the results can be improved with further experiments and trial-and-error method until a significant improvement in accuracy is achieved. The third layer of the model is the output layer, which consists of two neurons to classify the malignant and benign classes.

4. Proposed Model for chronic kidney disease prediction

The chronic kidney disease prediction model, which uses both LSTM and random forest algorithms, has undergone some adjustments to its parameters to achieve better results and increase accuracy. The general structure of the kidney disease prediction model is shown in Figure 2, which also consists of three layers: Data, Prediction, and Output. The data layer includes the kidney disease dataset and basic preprocessing steps similar to those of the breast cancer prediction model. In this study, we used existing data and specifically considered the chronic kidney disease dataset.

The prediction level settings were changed according to the specific requirements of the new data. The number of neurons was changed, while the main structure of the LSTM remains the same as that of the breast cancer model. The third layer of the chronic kidney disease model is the output layer, which contains two neurons for classifying "chronic kidney disease (CKD)" and "non-chronic kidney disease (NON-CKD)" classes.

5. Result

The dataset used in the study contains 400 records, which were divided into a training dataset with 280 rows and a test dataset with 120 rows. The balance of the data shows that there are 174 positive instances (ckd) and 106 negative instances (nckd). As can be seen in Figure 3, the data set also has some notable outliers, particularly in the bgr and br variables, which need to be addressed. In addition, the dataset contains missing values that also need to be addressed. To ensure the accuracy of the model, the outliers and missing values were treated before training the model. Previous research has shown that outliers can significantly affect the accuracy of algorithms. The results of the detailed analysis showed that machine learning algorithms are well suited for problems such as chronic kidney disease. While deep learning algorithms are effective for large data sets, logistic regression and random forest algorithms have been shown to be successful for smaller data sets.

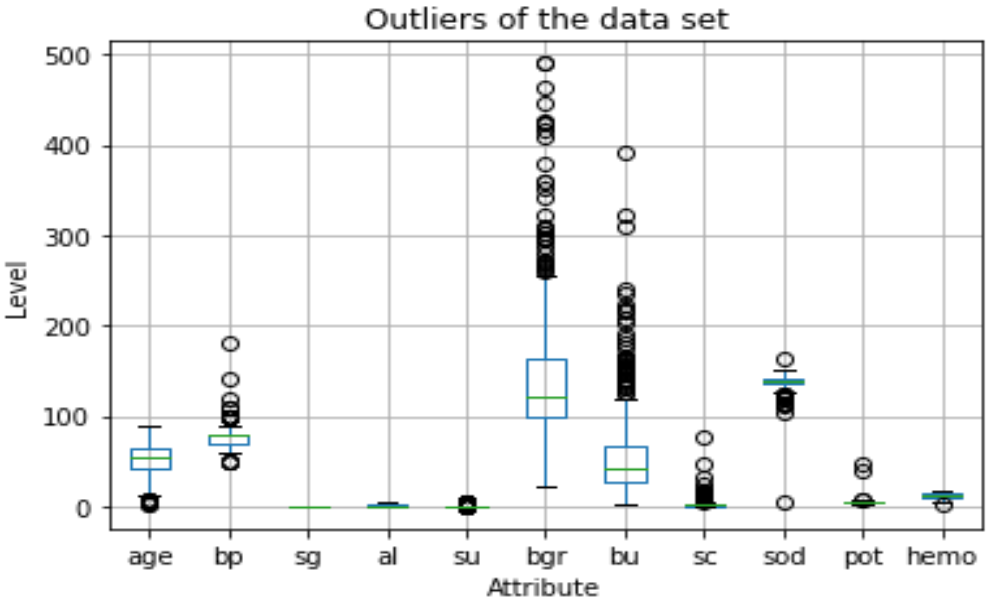

Figure 3: Outliers of the dataset.

6. Conclusion

In summary, our proposed multilevel machine learning approach is an effective way to improve the early detection and diagnosis of CKD and breast cancer. By combining supervised and unsupervised techniques and feature selection, the model can extract important information from the data and improve its prediction accuracy. The results of this study suggest that this hybrid predictive model has the potential to be used in clinical practice to improve patient outcomes. However, the model still needs to be evaluated in large-scale clinical trials to confirm its generalizability and robustness.

References

[1] Malakar, S., Roy, S. D., Das, S., Sen, S., Velásquez, J. D., & Sarkar, R. (2022). Computer Based Diagnosis of Some Chronic Diseases: A Medical Journey of the Last Two Decades. *Archives of computational methods in engineering : state of the art reviews*, 29(7), 5525–5567. https://doi.org/10.1007/s11831-022-09776-x

[2] Uddin, S., Khan, A., Hossain, M. et al. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak* 19, 281 (2019). https://doi.org/10.1186/s12911-019-1004-8

[3] Charilaou, P., & Battat, R. (2022). Machine learning models and over-fitting considerations. *World journal of gastroenterology*, 28(5), 605–607. https://doi.org/10.3748/wjg.v28.i5.605

[4] Alfred, R., & Obit, J. H. (2021). The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review. *Heliyon*, 7(6), e07371. https://doi.org/10.1016/j.heliyon.2021.e07371

[5] Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2022). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of ambient intelligence and humanized computing*, *1–28*. Advance online publication. https://doi.org/10.1007/s12652-021-03612-z

[6] Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. *Informatics in Medicine Unlocked*, 20, 100391.

[7] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*, 5, 8869-8879.

[8] Poonia, R. C., Gupta, M. K., Abunadi, I., Albraikan, A. A., Al-Wesabi, F. N., & Hamza, M. A. (2022). Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease. *Healthcare*.

[9] Zhang, D., Zou, L., Zhou, X., & He, F. (2018). Integrating feature selection and feature extraction

methods with deep learning to predict clinical outcome of breast cancer. *IEEE Access*, 6, 28936-28944.

[10] Raghupathi, V., & Raghupathi, W. (2017). Preventive healthcare: A neural network analysis of behavioral habits and chronic diseases. *Healthcare*.

[11] Malathi, D., Logesh, R., Subramaniyaswamy, V., Vijayakumar, V., & Sangaiah, A. K. (2019). Hybrid reasoning-based privacy-aware disease prediction support system. *Computers & Electrical Engineering*, 73, 114-127.

[12] Chen, G., Ding, C., Li, Y., Hu, X., Li, X., Ren, L., Ding, X., Tian, P., & Xue, W. (2020). Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the internet of medical things platform. *IEEE Access*, 8, 100497-100508.

[13] Torrisi, M., Pollastri, G., & Le, Q. (2020). Deep learning methods in protein structure prediction. *Computational and Structural Biotechnology Journal*, 18, 1301-1310.

[14] Walsh, S. L., Humphries, S. M., Wells, A. U., & Brown, K. K. (2020). Imaging research in fibrotic lung disease; applying deep learning to unsolved problems. *The Lancet Respiratory Medicine*, 8(11), 1144-1153.

[15] Pournaghi, S. M., Bayat, M., & Farjami, Y. (2020). MedSBA: a novel and secure scheme to share medical data based on blockchain technology and attribute-based encryption. *Journal of Ambient Intelligence and Humanized Computing*, 11(11), 4613-4641.