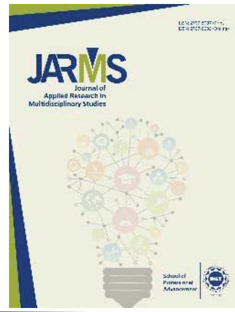
Journal of Applied Research and Multidisciplinary Studies (JARMS)
Volume 3 Issue 1, Spring 2022
ISSN(P): 2707-5087 ISSN(E): 2707-5095
Homepage: https://journals.umt.edu.pk/index.php/jarms



Article QR

Title:	**CLASSIFICATION APPROACHES FOR MULTI LEVEL BREAST CANCER PREDICTION: MACHINE LEARNING PERSPECTIVE**
Author (s):	Hujun Sun[1], Ling Weay Ang[2], Sellappan Palaniappan[3]
Affiliation (s):	Malaysia University of Science and Technology (MUST), Selangor, Malaysia
DOI:	https://doi.org/10.32350/jarms.31.03
History:	Received: February 8, 2022, Revised: March 1, 2022, Accepted: June 9, 2022
Citation:	Sun, H., Ang, L. W., & Palaniappan, S. (2022). Classification approaches for multi level breast cancer prediction: machine learning perspectives. *Journal of Applied Research and Multidisciplinary Studies*, *3*(1), 00–00. https://doi.org/10.32350/jarms.31.03
Copyright:	© The Authors
Licensing:	This article is open access and is distributed under the terms of Creative Commons Attribution 4.0 International License
Conflict of Interest:	Author(s) declared no conflict of interest

A publication of
School of Professional Advancements
University of Management and Technology, Lahore, Pakistan

Review of Multi-Level Prediction and Classification of Breast Cancer Based On Deep Learning and Machine Learning

Hujun Sun[1*], Ling Weay Ang[2] and Sellappan Palaniappan[3]

Malaysia University of Science and Technology (MUST)

Abstract

Breast cancer is the most common cancer in women worldwide and the leading cause of mortality in women as well. Numerous studies have been conducted on the diagnosis and prognosis of breast cancer. However, each technique carries a different accuracy rate depending on the situation, the tools and the data sets used. In the current study, the results of machine learning and deep learning algorithms for breast cancer prediction were compared. The comparison of results was carried out to determine the best strategy in order to manage large datasets, while maintaining good prediction accuracy. Various machine intelligence techniques including machine learning (ML) and deep learning (DL) were investigated in the context of breast cancer. The current paper also attempted to discuss some of the previous research work on machine learning algorithms, segmentation, and classification approaches used for multilevel prediction and classification of breast cancer. According to the results of the comparative research in this study, multiple disease prediction using recurrent neural networks (RNN) with long short-term memory (LSTM) leads to better classification performance in breast cancer prediction.

Keywords: breast cancer, deep learning (DL), long short-term memory (LSTM) machine learning (ML), recurrent neural networks (RNN)

Introduction

Chronic diseases contribute significantly to morbidity, mortality, disability, and low quality of life (Monzani & Pizzoli, 2020). According to WHO, the risk of dying from chronic diseases ranges from 56 to 80 per 1000 people. Cancer is another chronic disease that is said to be claiming a large number of lives in current times (Krishna et al., 2020). Numerous cancer prediction algorithms were developed using machine learning techniques; however, some of these algorithms lacked predictive accuracy. Early detection of cancer helps to save a significant number of lives. In recent years, clinicians

[*] Corresponding Author: sun.hujun@phd.must.edu.my

used time-consuming and error-prone manual techniques to detect cancer for instance, breast cancer, which is most common in women (Alahe & Maniruzzaman, 2021).

In recent years, artificial intelligence (AI), machine learning (ML), and deep learning (DL) techniques contributed to the automation of medical diagnostic work (Abd El Kader et al., 2021; Davenport et al., 2019; Ahuja, 2019). Researchers used prediction, classification, and segmentation techniques, such as Deep Convolutional Neural Networks (CNNs) to classify tumors in medical images, achieving higher accuracy (Amin et al., 2022; Gao, 2020). Several prediction algorithms were used to detect cancer in medical patients' data. For instance, Support Vector Machine (SVM), Neural Network (NN), LR, Nevin Biases (NB), Fuzzy Logic, Transfer Learning, Ensemble Learning, Transduction Learning, KNN, and Adaboost are most commonly used in various papers (McKinney et al., 2022). Additionally, Nave Bayes, Random Forest, simple logical regression, decision tree, linear regression model, and simple linear regression were also used to predict chronic diseases (Islam et al., 2021). Researchers also used decision tree algorithms for instance, the J48 algorithm to predict chronic kidney disease (CKD), which was shown to be quite efficient as it allows detection of all stages of the disease (Devi et al., 2021).

Traditionally, specialists diagnose chronic diseases using data from medical records and frequently asked questions from patients. Medical records provide a good picture of patients; however, their data would still be inaccurate as they aren't good enough to talk to doctors about their symptoms, which makes it difficult for the doctors to diagnose them (Grünloh, 2018). The biggest challenge is the diagnosis of chronic diseases due to certain limitations, such as lack of information about the disease symptoms. Patients aren't aware of their disease and may not be able to provide information about the symptoms or location of the disease. Resultantly, physicians find it difficult to predict the nature and duration of the disease, leading to inaccurate diagnosis and treatment (Malathi et al., 2019). Therefore, the aim of the current study was to investigate recently published segmentation and classification algorithms for breast cancer prediction and classification at multiple levels.

The current research attempted to explore the use of machine learning and deep learning algorithms for breast cancer prediction. It also focused on the comparison of their accuracy to determine the best strategy in order to manage large data sets while maintaining good prediction accuracy. Moreover, the study discussed some of the previous research on machine learning algorithms, segmentation, and classification approaches used for multilevel breast cancer prediction and classification. Afterwards, it presented the results of comparative research which suggested that multiple disease prediction using recurrent long-term memory (LSTM) neural networks leads to better classification performance in breast cancer prediction.

The current research may have significant impact on the field of breast cancer diagnosis and prognosis, as accurate prediction of breast cancer risk may lead to earlier detection and more effective treatment options. Additionally, the use of machine learning and deep learning algorithms may help process large amounts of data more efficiently and accurately, allowing for better predictive accuracy.

However, the study also observed some limitations. Firstly, it only compared the accuracy of machine learning and deep learning algorithms in breast cancer prediction. Moreover, other factors, such as cost, time, and ethical considerations were not considered. Secondly, the results of the study may not be generalizable to all populations, as the accuracy of machine learning and deep learning algorithms may vary depending on the demographic characteristics of the population studied. Finally, it only examined the use of recurrent long-term memory (LSTM), neural networks, and did not compare the accuracy of other machine learning and deep learning algorithms.

Literature Review

Machine Learning Approaches in Chronic Diseases Prediction

Chronic diseases are prohibitively expensive to treat, yet they are the leading cause of death globally (Lehnert et al., 2011). Due to the complexity of chronic diseases, clinicians are not able to diagnose it accurately in advance (Miotto et al., 2016). Manual diagnosis by clinicians using medical records is possible with a limited number of documents; however, diagnosis becomes more difficult when a more extensive record is created due to lack of time. It is possible that it may take longer to get the final results (Sun, 2018). Each chronic disease has a different variety of symptoms. Therefore, depending on the symptoms, a prediction for chronic diseases may be made using artificial intelligence (AI) and machine learning algorithms. The prediction algorithms provide accurate results in a short time which saves doctors' time. The Big Data analytics models have various applications in the medical field too. For instance, predictive models to mitigate the effects of chronic diseases, such as brain stroke (Raghupathi, 2014). The accuracy of predictive models is improved by proper feature extraction from the dataset. Machine learning models, such as Naive Bayes, K-Nearest Neighbor, and Decision Tree may classify the chronic diseases using multiple performance error metrics (Mahmood & Abdulazeez, 2021). Neural networks were applied in various data mining fields using back-propagation networks for pattern recognition problems. Multilayer perceptron’s also perform well in chronic disease prediction when the values of learning rate and training test data vary (Yildirim, 2017).

Deep learning techniques are essential in the medical field as they may be used to detect and evaluate a wide range of diseases including lung cancer (Gu et al., 2015) and breast cancer (Ostrom et al., 2014). Although, deep learning and neural network-based algorithms gained tremendous progress to categorize and predict the chronic diseases, overfitting remains a problem (Srivastava, 2014). If enough training data is available, deep learning methods may be used to solve the multi-label classification problem in chronic disease prediction (Zhang, 2020). If sufficient data is not available, other machine learning algorithms, such as k-NN, decision trees, and support vector machines (SVM) may prove helpful. Multi-label classification is a type of supervised learning used in a variety of applications including disease prediction, image classification, semantic analysis, and many others (Lin et al., 2021). Before applying the classification method, the algorithms usually convert the multi-label classification into binary classification. In order to transform multi-label classification into binary classification, decision trees, random forests, and SVM usually use label transformation algorithms (Gu et al., 2015; Murphy, 2012; Srivastava et al., 2014; Yoo et al., 2020).

Breast Cancer Diseases Diagnosis and Prediction using Machine Learning Techniques

Several researchers conducted breast cancer research in recent years using a variety of datasets, including the SEER database, mammography images, the Wisconsin dataset, and datasets from other hospitals. Machine learning techniques included support vector machine (SVM), random forest, logistic regression, decision tree (C4.5), and K-Nearest Neighbors (KNN network). Islam et al. (2020) used machine learning techniques to conduct a comparative study on breast cancer prediction. An automated disease detection system assists medical personnel to diagnose the disease and provides a reliable, effective, and timely response, thereby reducing the risk of mortality. The Wisconsin breast cancer dataset was taken from the UCI Machine Learning Database, a well-known machine learning database. The ANNs showed accuracy, precision, and F1 values of 98.57%, 97.88%, and 0.9890, respectively.

The combination of a clustering method and an efficient probabilistic vector support machine resulted in a Wisconsin Breast Cancer Detection (WBCD) solution with a prediction rate of 99.10% (Osman, 2017).

Latchoumiet TP determined a classification value of 98.4 percent and presented a particle swarm optimization weighting for classification based on SSVM. Using the support vector machine, classification models for two commonly used breast cancer benchmark datasets, the best overall accuracy for breast cancer detection was achieved at 98.80% and 96.33%, respectively (Osareh et al., 2010).

Omondiagbe et al. (2019) used the Wisconsin Diagnostic Breast Cancer (WDBC) dataset to investigate the Support Vector Machines, Artificial Neural Networks, and Nave Bayes. In the current study, dimensionality reduction techniques and three popular ML algorithms were used to classify the malignant and benign tumours in the WDBC dataset. SVM-LDA and NN-LDA outperformed the other ML classifier models in simulations. SVM-LDA was preferred over NN-LDA because NN-LDA required more processing time. The current research provided an intelligent strategy for breast cancer diagnosis that combined linear discriminant analysis and supported vector machines. In the validation dataset, the selected strategy provided good and promising results. It achieved a classification accuracy of 98.82%, a sensitivity of 98.41%, a specificity of 99.07%, and an area under the Receiver Operating Characteristic Curve of 0.9994. Xie et al. (2019) used supervised and unsupervised deep convolutional neural networks to analyse breast cancer histopathology images. They showed that their experimental results were superior to those accessible in other studies they discovered. The Inception ResNet V2 network was better than the Inception V3 network to analyse histopathology images of breast cancer.

Support Vector Machine (SVM), KNearest Neighbor (K-NN), and Random Forest (RF) machine learning techniques were used to identify the medical images as cancerous or benign. The results showed that SVM achieved a high degree of accuracy of about 97%. A variety of new features were used for this algorithm , such as bagging and boosting in order to improve its efficiency. Additionally, deep learning with CNN achieved an accuracy level of over 98%. Medical diagnosis, based on the International Classification of Diseases (ICD) was aggregated at different levels for prediction to meet the needs of different stakeholders.

The effectiveness of the proposed method was confirmed by its application to two different hospital medical datasets. The first data set included 7,105 patients with 18,893 visits, and the second data set included 4,170 patients with 13,124 visits. The first study revealed a wide range of patient characteristics. The results showed that a recurrent neural network with long-short memory units can accurately predict diagnoses at different levels of aggregation. The method achieved exact match values of 98.90% and 95.12% for aggregation of three-digit ICD codes and 96.60% and 96.83% for aggregation of four-digit ICD codes, respectively. These results suggest that the model could be used to predict future disease risks for patients. In addition, the method could be used to guide hospital information systems and lead to long-term improvements in patient care.

In another study by Kim et al. (2019), a deep-learning model was used to predict chronic diseases in individual patients by treating data in each class as one word, even when a large portion of the input values were missing. The char-RNN method was used to categorize cases in the Korea National Health and Nutrition Examination Survey (KNHANES) VI as either normal or as having one of five chronic diseases, namely hypertension, stroke, angina, myocardial infarction, or diabetes mellitus. Results showed that the char-recurrent neural network outperformed other models, such as k-nearest neighbor, mode, and multiple imputation, by an average of 10%.

Hasan et al. (2021) compared the SVM with other algorithms, such as K-NN, C4.5, Nave Bayes (NB), Kmeans, EM, Partitioning Around Medoids (PAM), and Fuzzy C-means. They found that SVM outperformed these approaches with an accuracy up to 97%. Similarly, according to Fatima et al. (2020), the support vector machine, which is based on parallel computation may analyse numerous datasets simultaneously and carries the highest accuracy rate among two tools. It shows a lower error rate and computation time than the decision tree and random forest in Weka and Spark. They analysed the datasets used and investigated image pre-processing techniques along with the strengths and limitations of previous studies. Different imaging modalities were presented in detail, along with their performance measurements and results, challenges, and research directions. In a scientific study by Asri et al. (2016), breast cancer was predicted using support vector machine, random forest, logistic regression, decision tree (C4.5), and K-Nearest Neighbours (KNN). All other classifiers were outperformed by Support Vector Machine, which achieved the highest accuracy (97.2%). Saleh et al. (2022) described an optimized deep recurrent neural network model for breast cancer diagnosis based on recurrent neural networks and the Keras tuner optimization technique. Five standard models ML were compared with the optimized deep RNN. As compared with the other models, the results showed that the optimized RNN, with the univariate features obtained the best CV and test results. Dewangan et al. (2022) proposed a unique back-propagation-boosting-recurrent-Wienmed model with a hybrid Krill-Herd-African-Buffalo optimization technique in the Python environment to detect breast cancer at an early stage.

Multi-Disease Prediction using LSTM Recurrent Neural Networks

A neural network consists of neurons and feedback loops. In situations where the preceding and following inputs are interdependent, the RNN shows unique advantages. The procession of large amounts of data with delays and abnormal noise outperforms the most machine learning methods. During the training process, RNNs are susceptible to gradient explosion and fading along with the short-term memory errors. By building more than two layers, deep learning creates a multilayer hierarchical data representation, usually in the form of a neural network. Multiple layers with linear or non-linear activation functions are combined to form neural network models. Recurrent neural network (RNN) architecture is a type of neural network that is capable to handle both sequential and parallel data.

Another type of RNN, Bidirectional RNN (BRNN), is optimized to access the input sequences with well-defined beginnings and ends. Since RNNs could only use data from the previous context, Bi-RNN enables additional improvements. The Bi-RNN is capable to process the data that originates from two different sources. One RNN processes the sequence sequentially, while the other reverses the process. In a multi-label classification task in functional genomics, Zhang et al. (2021) used a back-propagation neural network to predict the labels. LSTM was also useful to evaluate the clinic diagnosis data approach on). Maxwell et al. (2017) used a deep neural network to classify the chronic diseases and increased classification accuracy on multiple labels. Deep learning-based methods, such as vector generation with a deep belief network by Liang et al. (2014) and general disease classification with SVM, remained superior to other methods. Long short-term memory (LSTM) was also shown to be effective in order to predict hospital mortality (Jin et al., 2018).

Tsoumakas and Vlahavas (2007) presented the random k-label sets (RAKEL) algorithm to categorize proteins, documents, and scenes with multiple labels. The prediction of chronic diseases is a priority for researchers because they are harmful and may lead to patient's death. Men et al. (2021) proposed a deep learning-based approach to predict multiple diseases. In order to solve temporal problems, the system uses a long short memory (LSTM) network and extends it with the help of two processes (time-based and attention-based). A real healthcare dataset is used to validate the performance of the approach. The results carry implications to assist the physicians in diagnosis through the use of smart technologies, and helps to improve the overall quality of healthcare services too. The proposed method may be used to support clinical decision making. Wang et al. (2020) presented a method for the systematic assessment of future disease risks based on longitudinal medical data of patients. Medical diagnosis based on the International Classification of Diseases (ICD) was aggregated at different levels for prediction in the current study to meet the needs of different stakeholders.

Methodology

The literature revealed that most chronic disease prediction methods used machine learning algorithms for disease classification and prediction. Very few authors used neural networks and deep learning for chronic disease prediction. The proposed method was experimental in nature and used Python platform to implement the algorithms, such as deep learning, long short memory (LSTM), and Random Forest. LSTM was selected due to its strong predictive ability in the deep learning models. LSTM was an artificial recurrent neural network (RNN).

Therefore, the training time was faster than the other models. It was successfully used for image processing, speech recognition, and handwriting recognition. The support vector machine may also be a good choice for chronic disease prediction. Support vector machine (SVM) was used as a supervised learning model for classification problems. They possess a strong ability to predict a limited training dataset. Therefore, SVM was used for both linear and nonlinear classification problems. It carries a solid ability to categorize the unsupervised data based on statistical methods.

Discussion

Although, chronic disease prediction and classification achieved considerable progress, many elements of classification and prediction still need to be addressed. The proposed models displayed limitations, lack of effective feature extraction, and low prediction accuracy. Overfitting and underfitting are two drawbacks of models based on deep learning and neural networks. By including a feature extraction module, the problems associated with neural network models may be avoided. LSTM models in recurrent neural networks are ideal to process time series data that could capture long-term dependencies and nonlinear dynamics. As compared to other time series models, such as Markov models, conditional random fields do not perform well in learning dependencies while processing time series data. Therefore, expert knowledge or assistance in future engineering is required. Random discoveries occur infrequently; however, the LSTM is capable of doing so.

The LSTM model is designed to help physicians predict disease, based on the symptoms of new patients, saving time, and valuable resources for proper treatment along with further medical testing (Ker et al., 2017). Datasets were acquired from Kagle for breast cancer (U. M. Learning) and chronic kidney disease (C.-l. Lounge). The proposed model used long-term memory (LSTM) to predict breast cancer. The main focus of the model was on feature extraction in order to improve the accuracy as compared to the methods already proposed in the literature. A comparative analysis of these algorithms in terms of accuracy and statistical features was performed to find the most suitable algorithms for chronic disease prediction.

The literature revealed that most of the chronic disease prediction methods use machine learning algorithms for disease classification and prediction. LSTM was selected because of its strong predictive ability in deep learning models. LSTM is an artificial recurrent neural network (RNN). Therefore, the training time was faster than the other models. It was successfully used for image processing, speech recognition, and handwriting recognition. The support vector machine may also be a good choice for chronic disease prediction. It was used as a supervised learning model for classification problems. They carried strong prediction ability with limited training dataset. Therefore, SVM was used for both linear and nonlinear classification problems. It has a solid ability to categorize unsupervised data based on statistical methods.

The Random Forest was used for classification problems, where the algorithm-built decision trees based on the training data. The final class was determined based on the results of the trees. The proposed model was a mixture of these algorithms and therefore used the best algorithm based on the data. The selection of multiple algorithms was used to compare and analyse appropriate algorithms for each type of data. The prediction accuracy matrices were used to compare the accuracy of deep learning with machine learning algorithms for performance measurement. The proposed model represented a chronic disease prediction system using data from patients with breast cancer. Data from already available datasets was used for the experiments. Thus, no data collection was performed. The moving average method was used to clean and pre-process the data along with handling the missing values. Statistical methods were used to select the best features for analysis, since prediction accuracy depends on feature selection. Large data sets were used to develop prediction algorithms for breast cancer survival using two well-known data mining approaches, artificial neural networks (ANN), and decision trees (DT), along with a conventional statistical method.

Conclusion

The current paper presented a comprehensive overview of disease prediction methods in conjunction with machine learning (ML) algorithms, deep learning (DL), and neural network-based approaches. In the healthcare domain, the LSTM prediction model helps to improve the overall health of patients through rapid disease classification and prediction. It reduces the probability of chronic diseases. Hospitals may use the proposed predictive model for early prediction of chronic diseases, saving time, and resources if the model is appropriately evaluated and analyzed using patient data. Various healthcare stakeholders may use the prediction tool.

References

Abd El Kader, I., Xu, G., Shuai, Z., Saminu, S., Javaid, I., Ahmad, I. S., & Kamhi, S. (2021). Brain tumour detection and classification on MR images by a deep wavelet auto-encoder model. *Diagnostics (Basel)*, *11*(9), Article e1589. https://doi.org/10.3390/diagnostics11091589

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. *PeerJ*, *7*, Article e7702. https://doi.org/10.7717/peerj.7702

Alahe, M. A., & Maniruzzaman, Md. (2021). *Detection and diagnosis of Breast Cancer using deep learning* (Paper presentation). 2021 IEEE Region 10 Symposium (pp. 1–7). https://doi.org/10.1109/TENSYMP52854.2021.9550975

Amin, J., Anjum, M. A., Gul, N., & Sharif, M. (2022). A secure two-qubit quantum model for segmentation and classification of brain tumor using MRI images based on blockchain. *Neural Computing and Applications*, *34*(20), 17315–17328. https://doi.org/10.1007/s00521-022-07388-x

Asri, H., Mousannif, H., Al Moatassime, H., & Noël, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, *83*, 1064–1069. https://doi.org/10.1016/j.procs.2016.04.224

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. https://doi.org/10.7861%2Ffuturehosp.6-2-94

Devi, C. A., Abdul Jabbar, F., Varshini, S. K., Rithanya, K. M. M., & Naveena, K. S. (2021). Risks of chronic kidney disease prediction using various data mining algorithms. *International Journal of Informatics, Information System and Computer Engineering (INJIISCOM)*, *2*(2), 53–65. https://doi.org/10.34010/injiiscom.v2i2.6907

Dewangan, K. K., Dewangan, D. K., Sahu, S. P., & Janghel, R. (2022). Breast cancer diagnosis in an early stage using novel deep learning with hybrid optimization technique. *Multimedia Tools and*

Applications, *81*, 13935–13960. https://doi.org/10.1007/s11042-022-12385-2

Fatima, N., Liu, L., Sha, H., & Ahmed, H. (2020). Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. *IEEE Access*, *8*, 150360–150376. https://doi.org/10.1109/ACCESS.2020.3016715

Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., Wang, Y., Fu, H., & Dai, J. (2020). Mental health problems and social media exposure during COVID-19 outbreak. *PLoS ONE*, *15*(4), Article e0231924. https://doi.org/10.1371/journal.pone.0231924

Grünloh, C., Myreteg, G., Cajander, Å., & Rexhepi, H. (2018). "Why do they need to check me?" Patient participation through eHealth and the doctor-patient relationship: Qualitative study. *Journal of Medical Internet Research*, *20*(1), Article e11. https://doi.org/10.2196/jmir.8444

Gu, B., Sheng, V. S., & Li, S. (2015). *Bi-parameter space partition for cost-sensitive SVM* (Paper presentation). 24th International Joint Conference on Artificial Intelligence. Argentina.

Hasan, S., Sagheer, A., & Veisi, H. (2021). Breast Cancer classification using machine learning techniques: A review. *Turkish Journal of Computer and Mathematics Education*, *12*(14), 1970–1979.

Islam, Md. A., Akter, S., Hossen, Md. S., Keya, S. A., & Afrin, S., & Hossain, S. (2021). *Risk factor prediction of chronic kidney disease based on machine learning algorithms* (Paper presentation). 3rd International Conference on Intelligent Sustainable Systems. India. http://dx.doi.org/10.1109/ICISS49785.2020.9315878

Islam, Md., Haque, Md. R., Iqbal, H., Hasan, Md. M., Hasan, M., & Kabir, M. N. (2020). Breast Cancer prediction: A comparative study using machine learning techniques. *SN Computer Science*, *1*, Article e290. https://doi.org/10.1007/s42979-020-00305-w

Jin, M., Bahadori, M. T., Colak, A., Bhatia, P., Celikkaya, B., Bhakta, R., Senthivel, S., Khalilia, M., Navarro, D., Zhang, B., Doman, T., Ravi, A., Liger, M., & Kass-hout, T. (2018). Improving hospital mortality prediction with medical named entities and multimodal learning. *arXiv*, Article e1811.12276. https://doi.org/10.48550/arXiv.1811.12276

Ker, J., Wang, L., Rao, J., & Lim, T. (2017). Deep learning applications in medical image analysis. *IEEE Access*, *6*, 9375–9389. https://doi.org/10.1109/ACCESS.2017.2788044

Kim, C., Son, Y., & Youm, S. (2019). Chronic disease prediction using character-recurrent neural network in the presence of missing information. *Applied Sciences*, *9*(10), Article e2170. https://doi.org/10.3390/app9102170

Krishna, S. M., Omer, S. M., Li, J., Morton, S. K., Jose, R. J., & Golledge, J. (2020). Development of a two-stage limb ischemia model to better simulate human peripheral artery disease. *Scientific Reports*, *10*(1), 1–16. https://doi.org/10.1038/s41598-020-60352-4

Lehnert, T., Heider, D., Leicht, H., Heinrich, S., Corrieri, S., Luppa, M., Riedel-Heller, S., & Konig, H. H. (2011). Health care utilization and costs of elderly persons with multiple chronic conditions. *Medical Care Research and Review*, *68*(4), 387–420. https://doi.org/10.1177/1077558711399580

Liang, Z., Zhang, G., Huang, J. X., & Hu. Q. V. (2014). *Deep learning for healthcare decision making with EMRs* (Paper presentation). 2014 IEEE International Conference on Bioinformatics and Biomedicine (pp. 556-559). http://doi.ieeecomputersociety.org/10.1109/BIBM.2014.6999219

Lin, J., Cai, Q., & Lin, M. (2021). Multi-label classification of fundus images with graph convolutional network and self-supervised learning. *IEEE Signal Processing Letters*, *28*, 454–458. https://doi.org/10.1109/LSP.2021.3057548

Mahmood, I., & Abdulazeez, A. M. (2021). The role of machine learning algorithms for diagnosing diseases. *Journal of Applied Science and Technology Trends*, *2*(1), 10–19. http://dx.doi.org/10.38094/jastt20179

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. https://doi.org/10.1016/j.compeleceng.2018.11.009

Maxwell, A., Li, R., Yang, B., Ou, A., Hong, H., Zhou, Z., Gong, P., & Zhang, C. (2017). Deep learning architectures for multi-label classification of intelligent health risk prediction. *BMC Bioinformatics*, *18*(14), 121–131. https://doi.org/10.1186/s12859-017-1898-z

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Gracia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., . . . Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*, 89–94. https://doi.org/10.1038/s41586-019-1799-6

Men, L., Ilk, N., Tang, X., & Liu, Y. (2021). Multi-disease prediction using LSTM recurrent neural networks. *Expert Systems with Applications*, *177*, Article e14905. https://doi.org/10.1016/j.eswa.2021.114905

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, *6*(1), 1–10. https://doi.org/10.1038/srep26094

Monzani, D., & Pizzoli, S. F. M. (2020). The prevention of chronic diseases through eHealth: A practical overview. In P. Gabriella & T. Stefano (Eds.), *P5 eHealth: An agenda for the health technologies of the future* (pp. 33-51). Springer Nature.

Murphy, K. P. (2012), *Machine learning: A probabilistic perspective*. MIT press.

Omondiagbe, D. A., Veeramani, S., & Sidhu, A. S. (2019). Machine learning classification techniques for Breast Cancer diagnosis. *Material Science and Engineering*, *495*, Article e012033 https://doi.org/10.1088/1757-899X/495/1/012033

Osareh, A., & Shadgar, B. (2010). *Machine learning techniques to diagnose breast cancer* (Paper presentation). 5th International Symposium on Health Informatics and Bioinformatics (pp. 114 – 120). https://doi.org/10.1109/HIBIT.2010.5478895

Osman, A. H. (2017). An enhanced breast cancer diagnosis scheme based on two-step-SVM technique. *International Journal of Advanced Computer Science and Applications*, *8*(4), 158–165. https://dx.doi.org/10.14569/IJACSA.2017.080423

Ostrom, Q. T., Gittleman, H., de Blank, P. M., Finlay, J. L., Gurney, J. G., McKean Cowdin, R., Stearns, D. S., Wolff, J. E., Liu, M., Wolinsky, Y., Kruchko, C., & Barnholtz-Sloan, J. S. (2016). American brain tumor association adolescent and young adult primary brain and central nervous system tumors diagnosed in the United States in 2008-2012. *Neuro-Oncology*, *18*(Suppl 1), i1–i50. https://doi.org/10.1093/neuonc/nov297

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, *25* (1-2), 204–216. https://doi.org/10.1186/2047-2501-2-3

Saleh, H., Abd-El Ghany, S. F., Alyami, H., Alosaimi, W. (2022). Predicting breast cancer based on optimized deep learning approach. *Computational Intelligence and Neuroscience*, *2022*, Article e1820777. https://doi.org/10.1155/2022/1820777

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Data processing and text mining technologies on electronic medical records: A review. *Journal of Healthcare Engineering*, *2018*, Article e4302425. https://doi.org/10.1155/2018/4302425

Tsoumakas, G., & Vlahavas, I. (2007). *Random k-labelsets: An ensemble method for multilabel classification* (Paper presentation). European Conference on Machine Learning (pp. 406–417). https://doi.org/10.1007/978-3-540-74958-5_38

Wang, T., Qiu, R. G., Yu, M., & Zhang, R. (2020). Directed disease networks to facilitate multiple-disease risk assessment modelling. *Decision Support Systems*, *129*, Article e 113171. https://doi.org/10.1016/j.dss.2019.113171

Xie, J., Liu, R., Luttrell, J., & Zhang, C. (2019). Deep learning-based analysis of histopathological images of Breast Cancer. *Frontiers in Genetics*, *10*, Article e80. https://doi.org/10.3389/fgene.2019.00080

Yildirim, P. (2017). *Chronic kidney disease prediction on imbalanced data by multilayer perceptron: Chronic kidney disease prediction* (Paper presentation). 2017 IEEE 41st Annual Computer Software and

Applications Conference (pp. 193–198). https://doi.org/10.1109/COMPSAC.2017.84

Yoo, S. H., Geng, H., Chui, T. L., Yu, S. K., Cho, D. C., Heo, J., Choi, M. S., Choi, I. H., Van, C. C., Nhung, N. V., Min, B. J., & Lee, H. (2020). Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Frontiers in Medicine*, *7*, Article e427. https://doi.org/10.3389/fmed.2020.00427

Zhang, J., Chen, L., Tian, J. X., Abid, F., Yang, W., & Tang, X. F. (2021). Breast cancer diagnosis using cluster-based under sampling and boosted C5.0 algorithm. *International Journal of Control, Automation and Systems*, *19*, 1998–2008. https://doi.org/10.1007/s12555-019-1061-x

Zhang, W., Zhao, Y., Zhang, F., Wang, Q., Li, T., Liu, Z., Wang, J., Qin, Y., Zhang, X., Yan, X., Zeng, X., & Zhang, S. (2020). The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The perspectives of clinical immunologists from China. *Clinical Immunology*, 214, Article e108393. https://doi.org/10.1016/j.clim.2020.108393