Volume: 5, Issue: 2
Page: 102-113
2021

International Journal of
Science and Business
Journal homepage: ijsab.com/ijsb

IJSAB International

Human Diseases Detection Based On Machine Learning Algorithms: A Review

Nareen O. M. Salim & Adnan Mohsin Abdulazeez

Abstract:
One of the most significant subjects of society is human healthcare. It is looking for the best one and robust disease diagnosis to get the care they need as soon as possible. Other fields, such as statistics and computer science, are needed for the health aspect of searching since this recognition is often complicated. The task of following new approaches is challenging these disciplines, moving beyond the conventional ones. The actual number of new techniques makes it possible to provide a broad overview that avoids particular aspects. To this end, we suggest a systematic analysis of human diseases related to machine learning. This research concentrates on existing techniques related to machine learning growth applied to the diagnosis of human illnesses in the medical field to discover exciting trends, make unimportant predictions, and help decision-making. This paper analyzes unique machine learning algorithms used for healthcare applications to create adequate decision support. This paper intends to reduce the research gap in creating a realistic decision support system for medical applications.

IJSB
Literature review
Accepted 19 January 2021
Published 25 January 2021
DOI: 10.5281/zenodo.4467510

Keywords: Human disease, Healthcare, Machine learning, Deep learning, Convolutional Neural Networks.

About Author (s)

Nareen O. M. Salim *Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.*
Adnan Mohsin Abdulazeez(corresponding author), *Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.* Email: nareen.mohameed@dpu.edu.krd

Introduction

In human society, healthcare is one of the most urgent issues, as the quality of life of people is It relies explicitly on it (Bagga & Hans, 2015). The healthcare area, however, is exceedingly varied, broadly dispersed, and fragmented. The delivery of adequate patient care from a clinical perspective requires access to appropriate patient information, rarely accessible when necessary (Grimson et al., 2001; Zeebaree et al., 2019a). Besides, the large variance in the order of tests for diagnostic purposes indicates the need for an adequate and suitable collection of tests (Daniels & Schroeder, 1977; Wennberg, 1984; Zebari et al., 2021; Zeebaree et al., 2019a). Smellie et al. (2002) expanded this claim by suggesting that the significant differences found in the request for general practice pathology arise primarily from individual variations in clinical practice and are thus likely to improve through more transparent and better-informed decision-making for physicians (Bargarai et al., 2020; Stuart et al., 2002). Therefore, medical data also consist of many heterogeneous variables obtained from various sources, such as demographics, history of illness, medications, allergies, biomarkers, medical photographs, or genetic markers, each offers a different partial view of the condition of the patient. Also, among the sources, as mentioned earlier, statistical properties are fundamentally different.

Researchers and practitioners face two challenges when analyzing such data: The curse of dimensionality (the number of dimensions and the number of samples increases exponentially in the space of the features) and the heterogeneity of function sources and statistical features (Pölsterl et al., 2016; Zebari et al., 2020; Zeebaree et al., 2019b). These causes contribute to delays and inaccuracies in the diagnosis of the disease and, therefore, patients have not been able to obtain adequate care. Therefore, there is a strong need for an appropriate and systematic approach that enables early detection of the disease and can be used as a physician's decision-making aid (Zhuang et al., 2009). Therefore, the medical, computer, and statistical fields face the challenge of exploring new strategies for modeling disease prognosis and diagnosis, as conventional paradigms struggle to answer all of this information (Huang et al., 2007). Today, ML offers many essential resources for intelligent data analysis. Furthermore, its technology is currently well adapted for the study of medical data. In particular, a wide variety of medical diagnostic work has been carried out on small-specialized diagnostic problems(Bargarai et al., 2020; Kononenko, 2001), where initial ML applications have been found. ML classifiers have been successfully used, for example, to differentiate between stable patients and those with Parkinson's disease (V.S. Sriram et al., 2016; R. Zebari et al., 2020), which is a valuable tool in clinical diagnosis. Indeed, on a wide range of significant issues, most ML algorithms perform very well.

2. Background Theory

This section briefly introduced Machine learning, its types, and the most used literature techniques, comparing the studies and research about machine learning.

2.1 ML

ML is a branch of artificial intelligence that enables computers to think like human beings and make their own decisions without human interference (Siraj et al., 2011; Kharya, 2012; Padmapriya & Velmurugan, 2015). ML has much progress in detecting various forms of disease due to the rapid growth of Artificial Intelligent. A machine learning algorithm also provides us with more precise predictions and performance (Shaheamlung et al., 2020). ML has been widely divided into various forms, as seen in figure 1. below.

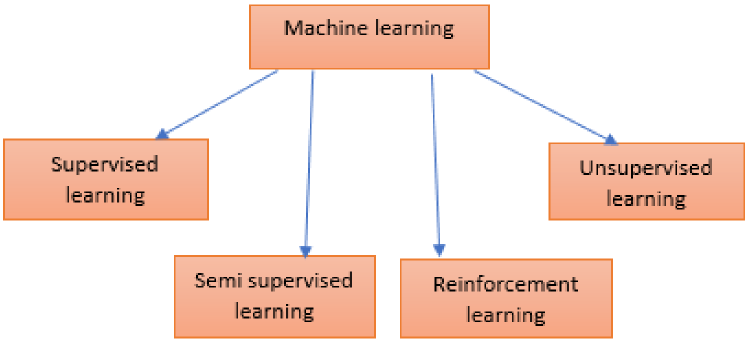

Figure 1: Different kinds of ML(Ayodele, 2012)

a) SUPERVISED LEARNING

This type of ML gives a training data set. This ML approach responds accurately to all feasible inputs, as it depends on the training data set. Supervised learning from examples is often referred to as learning (Hashem et al., 2018; Mahmood & Abdulazeez, 2019; Sadeeq & Abdulazeez, 2018; Shi & Malik, 2000; Zebari et al., 2020). Regression and classification are two forms of supervised machine learning.

b) UNSUPERVISED LEARNING

Right answers or goals are not given. Because of these similarities, the purpose of un-regulated learning techniques is to discover the similarities between knowledge data and the story structured by an un-directed learning approach. This type of learning is otherwise referred to as calculating thickness. Grouping requires unsupervised adaptation (Jahwar, 2021; Adeen et al., 2020; Pan & Tompkins, 1985).

c) SEMI SUPERVISED LEARNING

This method is known as the class of supervised learning techniques. This ML uses un-labeled data (Tajbakhsh et al., 2016) for training. Among controlled education and unsupervised learning, it is learning that occurs. Supervised learning has classified data, and unlabeled data is available for unsupervised learning.

d) REINFORCEMENT LEARNING

The psychology of behaviorists endorses this form of ML. An algorithm indicates that the answer is incorrect, but it does not say how to correct that response. This algorithm conducts several tests before it finds the right answer. Improvement is not feasible in this learning process(Hu et al., 2006).

2.2 Different techniques used by ML

Many scientists have developed all of the different machine learning algorithms to diagnose illnesses. The researcher states that ML operates efficiently to analyses various diseases. In the field of medicine, there are several levels of machine learning. Protein-protein collaboration used the observational space of programmed learning in some capabilities such as medical imaging, therapeutic knowledge retrieval, restorative choice assistance, and general patient administration. ML is used to differentiate and evaluate pneumonia, malignant growth of the lungs, and multiple ailments (Iswanto et al., 2019; Zebari et al., 2020; Zeebaree et al., 2020).

2.2.1. Support Vector Machines (SVM)

SVM, which was designed in the 1990s. SVM is used to accomplish (ML) tasks, and it is a prominent and straightforward tool. A selection of training samples divides each sample into different categories in this process. Help vector SVM computer, used primarily for problems with classification and regression (Abdulqader et al., 2020; Murphy, 2012; Zeebaree et al., 2019a).

2.2.2 k-nearest neighbors (k-NN)

One of the ML communities' well-known techniques is the k-NN classifier, described as a non-parametric approach (Al-Zebari & Sengur, 2019; Kramer, 2013). The k-NN classifier will consider training samples, a distance function, and several nearest neighbors (k). For distance measurement, Euclidean distance is a general solution. The class labels of the test specimens shall be decided by a majority vote of the predetermined labels of the k-nearest neighbors.

2.2.3 Logistic Regression (LR)

LR is defined as a generalized linear model. Two components, namely the linear component and the relationship function, consist of generalized linear models(Sadiq et al., 2020). The linear component of the classification model is calculated, and, through the relationship function, the output of this measurement is expressed. In the case of logistic regression, the linear outcome is run via a logistic function. Only values between 0.0 and 1.0 are returned by the logistic function (Kousarrizi et al., 2012; Sadeeq & Abdulazeez, 2018).

2.2.4 Decision trees (DTs)

The well-known basic non-parametric supervised machine learning methods used in data classification tasks (Al-Zebari & Sengur, 2019; Safavian & Landgrebe, 1991) are DTs. DTs' primary objective is to build a model that predicts a test sample's classmark by learning some rules that have been inferred from the training dataset. There are two types of nodes within a DT setup, such as leaf and internal nodes. Of the training examples entering the plate, a leaf has a classmark measured by the majority vote. Every internal node is a matter of operation, and it branches out according to the answers(Conference, 1999; Dey, 2016; Hota, 2014).

2.2.5 Naive Bayes classification

An example is the statistical classifiers for Bayesian classifiers. Naive Bayes determines class membership probabilities based on the classmark given (Hazra et al., 2016). It conducts one data scan, and therefore it is simple to classify.

2.2.6 Deep Learning (DL)

DL is concerned with knowledge processing using deep networks. It is an aspect of machine learning techniques. In its previous appearance in 1943, McCulloch and Pitts referred to DL as "cybernetics"(Mcculloch & Pitts, 1990). Researchers were attracted to DL because of its ability and characteristics to imitate the brain processes the way information before making decisions.

2.2.7 Convolutional Neural Networks (CNNs)

CNNs is an artificial neural network able to derive data from local characteristics. By assigning weights to accurate feature mapping, CNNs simplifies the network model, enabling total value reduction. The widespread use of CNNs in pattern recognition has resulted in these features (Huang & LeCun, 2006; Vincent et al., 2008).

2.2.8 Lasso Regression

Lasso regression is a type of shrinkage-using linear regression. Shrinkage is when, as they suggest, data values are reduced to a central point lasso strategy supports direct, sparse models (i.e., models with fewer parameters). This particular form of regression is ideal for models with high multicollinearity levels or when certain aspects of model selection are automated, such as variable selection/parameter elimination (Chitra, 2018).

2.3 Healthcare ML applications

In recognizing intricate patterns inside large and successful data, ML algorithms are useful. This facility is especially well-suited for clinical applications, particularly for people who rely on advanced genomics and proteomics measurements. It is also used in the diagnosis and detection of various diseases. ML algorithms can generate higher decisions on patients' treatment plans in medical applications by implementing sound health care systems (Mohammed et al., 2016; Rajabion et al., 2019). Hospital management uses this approach to forecast wait times for patients waiting for positions in the department of demand. These models use patient details, pain levels, demand department charts, and even the hospital room layout to infer wait times. Clinics can consider emergency room admissions using the predictive model. Thus, machine learning implementation may benefit patients by lowering costs, increasing precision, or disseminating short-term experience (Ahmad et al., 2018; Ramana et al., 2011).

3. Related work

There are many research areas and related works on this topic. In (Venkata Ramana et al., 2011), they found that the AP datasets were better than the UCLA datasets for all the various chosen algorithms. The writers used two separate datasets of inputs. The AP data sets were calculated to be better than the UCLA dataset. Based on the usefulness of their KNN classification, backward propagation and SVM give better outcomes. For the entire chosen algorithm, the AP data set is better than UCLA. Besides, 95.07, 96.27, 96.93, 97.47, & 97.07 % accuracy have C4.5, Backward propagation, Naïve Bayes, SVM, and KNN. (Kousarrizi et al., 2012) this analysis is focused on two databases on thyroid disease. The first dataset is taken from the UCI machine learning repository. The second is the actual data gathered from the Imam Khomeini hospital by the Intelligent Device Laboratory of the K.N.Toosi University of Technology. They obtained a classification accuracy of 98.62 % using SVM for the first dataset, which is the highest accuracy achieved so far. Chitra et al. (2018) showed the SVM with a Radial base function kernel is used for classification. The output parameters are high, such as the classification accuracy, sensitivity, and specificity of the SVM and RBF, making it the right choice for the classification process. Twelve morphological features from the ST segment were extracted (Fan et al., 2013). Using the SVM classifier, they obtained 95.20% sensitivity, 93.29% specificity, and 93.63% accuracy. Hariharan et al. (2014) diagnosed Parkinson's disease, in this approach, the neural networks and the SVM algorithm are fused. The experimental findings show that for Parkinson's dataset, the combination of feature preprocessing, feature reduction/selection methods, and classification give a maximum classification precision of 100 %. Senturk & Kara (2014) intended to contribute to early breast cancer diagnosis in this study. An analysis of the diagnosis of breast cancer for patients is provided. Seven different algorithms are used to realize the predictions of the other patients and give them precision. Patient data from UCI ML during the prediction process, the data mining tool RapidMiner 5.0, is used to apply data mining with the desired algorithms during the prediction process. In a difference between two classification algorithms, SVM and ANNs, was addressed by Vijayarani & Dhayanand (2015). In this study reached the target of predicting CKD based on their respective accuracies and timings. The one picked with higher accuracy, and the right timing was chosen. Survey of a paper conducted by Hashem et al. (2017) to classify liver disease. Different data mining classification methods were studied in this analysis, and the AP liver dataset data set used had better results than the UCLA dataset and concluded that C4.5 had achieved better results than other algorithms. Ko et al. (2017) used thermoscopic and clinical images that displayed the performance of CNNs approach, a CNNs architecture was trained from scratch. However, because of the limited datasets, a network's training from scratch to detect skin cancer is usually not viable. Most of the researchers, therefore either fine-tuned the model or used pre-trained models.

The experiments will be conducted on an experimental database. Based on classification accuracy obtained, three distinct characteristics, such as spectral, wavelet, and complexity-related characteristics, are computed and compared. Three distinct features, such as spectral, wavelet, and complexity-related features, are computed and compared based on classification accuracy obtained (Kulkarni & Bairagi, 2017). In a study (Acharya et al., 2017), researchers implemented a CNNs algorithm to detect regular and MI ECG beats (noise and noise). Using ECG beats with noise and noise reduction, they achieved an average precision of 93.53 % and 95.22 %. (Zeebaree et al., 2018) the writers explored a deep learning algorithm for microarray data classification based on the CNNs in the current research. CNN found that not all data had better performance than related techniques such as Elimination of Vector Machine Recursive Function and Enhanced Random Forest (SVM-RFE-iRF and varSeIRF). Most experimental studies on cancer datasets have shown that CNNs is superior in accuracy and gene minimization in cancer classification to hybrid mSVM-RFE-iRF. (Hashem et al., 2018) two algorithms, Backpropagation, SVM and the UCI system repository dataset, were used in this paper. Furthermore, SVM has an accuracy of 71 per cent higher than Backpropagation accuracy of 73.2 per cent for liver disease diagnosis. (Ahmed et al., 2019) Show that state-of-the-art techniques that take multimodal diagnosis into account have better accuracy than the manual diagnosis. The goals of this research attempt are as follows:1- Increase the accuracy levels comparable to state-of-the-art techniques; 2- To overcome the overfitting problem, 3- to examine proven brain landmarks that provide AD diagnosis with discernible features. First, authors integrate sets of simple CNNs as feature extractors and soft-max cross-entropy as the classifier to achieve the goals. After the preprocessing steps, they manually localized the left and right hippocampus and fed three-view patches to the CNNs. They have 90.05% precision. On the same dataset they used, the authors contrasted their model with the state-of-the-art methods and found our findings comparable. The efficiency analysis of the ML techniques on diabetes disease detection is performed in this paper. The work uses various ML techniques (DT), LR, DA, SVM, k-NN and ensemble learners. Software from MATLAB is taken into account. The findings are analyzed based on the 10-fold cross-validation criterion, and the performance analysis uses average classification accuracy. The average accuracy scores obtained are in the 65.5 % and 77.9 % range. The LR method provides the best accuracy score of 77.9 %, and the worst one is provided by the Coarse Gaussian SVM technique of 65.5 % in (Al-Zebari & Sengur, 2019). Durai (2019) analyzed the patient data sets based solely on a commonly diagnosed classification model for predicting the subject having a liver disorder. A necessary assessment process is carried out, depending on the studies, to maintain the integrity of a specific representation of the outcome. The J48 algorithm is a higher-performing algorithm with an accuracy rate of 95.04 per cent for feature selection. The output of tumour classification techniques for classifying MR brain image characteristics as n/a, gliomatosis, multifocal, and multicentric was analyzed (Cinarer & Emiroglu, 2019) study. KNN, RF, LDA and SVM machine learning algorithms tested these results. Compared to other algorithms, the SVM algorithm with a 90% precision rate was higher. Javeed et al. (2019) addressed overfitting, and developed a model to improve heart disease prediction; overfitting implies that the proposed model works and provides better data testing accuracy and gives unfortunate accuracy results for training data when predicting heart disease. They have built a model to solve this problem to give the best precision for training and testing results. There are two algorithms in the model: RAS (Random Search Algorithm) and the other is a random forest algorithm used for model prediction. In both training data and testing data, this proposed model provided them with better performance. Intracerebral hemorrhage sources for high mortality rate as a result, (Liu et al., 2019) it is based on multivariate analysis to anticipate the expansion of hematoma in spontaneous ICH with normally accessible SVM data and pointed out 83. A randomized 179 search approach was used in this study for parameter tuning, and recursive function 180 elimination was used for feature selection. Patient selection for thrombolytic procedures is another significant factor. In this study (Rustam et al., 2020), concerns use three types of the forecast for each model: the number of cases freshly infected, the number of casualties, and the number of recoveries over the next ten days. The outcomes provided by the Study Analysis indicate that the use of these methods in the current COVID-19 pandemic scenario is a promising mechanism. The results show that of all the models used, and the ES performs best, followed by LASSO & LR, which performs well in forecasting newly recorded incidents, death rate and recovery rate, Although SVM does not perform well in the prediction scenarios, the available dataset is given. Tanveer et al. (2020) analyzed 165 articles from 2005-2019 using different feature extraction techniques and machine learning techniques. Three key categories are studied in ML techniques: SVM, ANN and DL, and the ensemble methods.

Naqi et al. (2020) focused on 3D properties in the feature's extraction process. In image processing, recent developments in deep learning are a breakthrough. From traditional handcraft characteristics to deep automated characteristics, the emphasis of mechanical diagnostic systems has shifted. It helps in better identification and classification with a CT picture of nodular objects. For better feature reduction and type, an autoencoder and SoftMax are considered useful tools. By Kumar et al. (2020) employing DL techniques, namely CNNs, the proposed model eradicates errors in the manual process. The model, trained on cells' images, preprocesses the images first and extracts the best characteristics. This survey is followed by the optimized Dense Convolutional neural network structure (called DCNN) training the model and eventually predicting the type of cancer present in the cells. The model correctly replicated all measurements while accurately recollecting the samples 94 times out of 100. The aggregate accuracy was 97.2%, which is better than the techniques of CNNs such as SVMs, DT, RF, NB. This research shows that the DCNN model's performance is similar to that of the architectures of the developed CNNs with much fewer parameters and computation time tested on the retrieved dataset. Therefore, to evaluate the form of cancer in the bone marrow, the model can be used effectively.

Table 1. Machine learning Techniques for Diagnosis of Different Diseases.

Authors	Diseases	Dataset	Methods	Accuracy	Research Objective
(Kumar et al., 2020)	Blood Cancer	SN-AM	CNN	97.2%	By employing DL techniques, namely CNNs, the proposed model eradicates the manual method's likelihood of errors. The model, trained on cells' images, preprocesses the images first and extracts the best characteristics.
(Naqi et al., 2020)	Lung cancer	LIDC-IDRI (Meng et al., 2018)	DL	96.9%	Because the system's problem includes false-positive results, this work provides an automated detection system and classification to promote radiologists' diagnosis.
(Rustam et al., 2020)	Covid-19	GitHub (Wissel et al., 2020)	ES, LR, LASSO, SVM	------	The purpose of this research Provides displays the potential of ML models to estimate the number of future patients affected by COVID-19, which is widely regarded as a possible danger to humanity.
(Liu et al., 2019)	Brain stroke	1157 patients	SVM	83.3%	The expanding of hematoma is in anticipation that spontaneously ICH derives from accessible comparable by the usage of SVM
(Javeed et	Heart	Cleveland	RSA, RF	93.33%	Develop an intelligent system that would

al., 2019)	disease	heart failure (Meng et al., 2018)		(RSA+RF)	show good performance on both training and testing data diagnosis of heart failure.
(Cinarer & Emiroglu, 2019)	Brain tumour	(TCIA) (Scarpace et al. 2015)	KNN, RF, SVM and LDA	SVM: 90%	The best ML and classification algorithms' goal is to learn from training automatically and ultimately make a wise decision with high accuracy.
(Durai et al., 2019)	Liver disease	UCI (Shi & Malik, 2000)	J48, SVM& NB	With 95.04, the J48 algorithm has a better choice of features.	To predict the same definitive result, compare algorithm techniques with a higher accuracy rate for detecting liver disease.
(Ahmed et al., 2019)	Alzheimer Diseases	ADNI	CNN	90.05%	The study's objective is to increase the degree of accuracy comparable to state-of-the-art techniques, address the problem of overfitting, and examine validated brain technologies that include noticeable AD diagnostic features.
(Zeebaree et al., 2018)	Cancer disease	Different cancer dataset	CNN	100%	Based on gene expression data, DL algorithm applications are used to diagnose the disease.
(Acharya et al., 2017)	myocardial infarction	Control:40 CHD:7 (Pan & Tompkins, 1985; Singh & Tiwari, 2006)	CNN	98.99%	This study proposed diagnosing MI using 11 deep CNNs layers automatically, using two separate databases (noise and without noise).
(Kulkarni and Bairagi, 2017)	Alzheimer disease	100 (50 CN, 50 AD) (Kulkarni & Bairagi, 2017)	SVM	96%	The purpose of this research paper is to examine various characteristics of Alzheimer's disease diagnosis to serve as a potential biomarker to differentiate between the topic of AD and the ordinary subject.
(Senturk et al., 2014)	breast cancer	UCI	SVM, NB, KNN and DT	K-NN:95.15%, SVM:96.40%	Determine the best approaches to lead to early breast cancer detection. An overview of the diagnosis of breast cancer in patients is given.
(Hariharan et al., 2014)	Parkinson's disease	PD dataset was used from (UCI)	SVM	100%	found the best and an integrated approach to propose to improve the accuracy of detection of Parkinson's disease
(Kumari and Chitra, 2013)	Diabetic Disease	UCI	SVM	78%	Determine the best approaches to lead to early breast cancer detection. An overview of the diagnosis of breast cancer in patients is given.
(Kousarrizi et al., 2011)	Thyroid Disease	UCI	SVM	98.62%	Choose the best methods of feature selection and classification for thyroid disease diagnosis, which is one of the most critical classification problems

Discussion

This paper discusses various instruments and methods commonly used in the fields of medicine and healthcare. These tools are within ML and allow us to reach DL's main aim, finding useful patterns in databases, explaining and making a non-trivial prediction about data. We summarized the technical details shown in table 1: (including the References, Year, Diseases, Dataset, Performance and Research Objective) of the research mentioned in this previous section. As shown in table 1: some researchers used DL algorithms to achieve a higher rate of deeper detecting to improve precision, trust, and performance. It has been noticed that five researchers (Kumar et al., 2020; Naqi et al., 2020; Ahmed et al., 2019;

Zeebaree et al., 2018, and Acharya et al., 2017). Focused on the DL algorithms for a detect disease like (Blood cancer, Lung cancer, Alzheimer, Cancer disease and myocardial infarction) show the performance column the accuracy of CNNs in cancer disease has a higher rate than the others disease. Classification is the model used to search for a model or function that defines and distinguishes the data, classes, or concepts that the model uses to predict the class of object whose class mark is unknown. In classification, they create software that can learn how the data objects can be categorized. The derived model can be presented as classification or rules; many researchers have used different algorithms to help health care practitioners diagnose diseases with greater precision in diagnosis. In this study many classification algorithms used for detect disease as (LR, LASSO, SVM, KNN, RF, LDA, NB, J48, RSA and DT) as shown in table 1, SVM in (Liu et al., 2019; Cinarer & Emiroglu, 2019; Kulkarni and Bairagi, 2017; Senturk et al.,2014; Hariharan et al., 2014; Kumari and Chitra, 2013, and Kousarrizi et al., 2011) has the higher accuracy among the other classification algorithms for the disease detection. However, given the available dataset, the Rustam et al., (2020) SVM performs poorly in all prediction scenarios and the Durai et al. (2019) J48 algorithm is considered a better output algorithm when it comes to feature selection with an accuracy rate of 95.04 %.

Conclusion

Intelligent data processing is a social necessity for identifying, as soon as possible, of useful and robust disease detections to provide patients with appropriate care within the shortest possible time. This detection has been carried out in recent decades by detecting exciting patterns in databases. Smart data processing is emerging as a requirement for effective and robust diseases to be found by society. Detection of patients providing the necessary treatment as soon as possible within the shortest possible period. This identification has been achieved in recent decades through the method of identifying exciting patterns in databases. A comprehensive overview of intelligent data analysis tools in the medical sector is given in this paper. Some examples of some algorithms used in these medical field areas are also presented, examining potential patterns based on the target searched, the methodology used, and the application field. Given the pace at which new works emerge in this emerging field, a systematic analysis such as the one we have just presented may become obsolete in a short period. For this reason, we consider that, after a careful quest for new scientific literature, Table 1 should mainly be revised, provided that further research is more likely to take place in the short term on the application of established techniques in this field than on the proposal of new techniques which are novel and not merely enhancing or changing existing ones.

References

Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Machine learning supervised algorithms of gene selection: A review. *Technology Reports of Kansai University*, *62*(3), 233–244.

Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*, *415–416*, 190–198. https://doi.org/10.1016/j.ins.2017.06.027

Ahmad, M. A., Teredesai, A., & Eckert, C. (2018). Interpretable machine learning in healthcare. *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018*, 447. https://doi.org/10.1109/ICHI.2018.00095

Ahmed, S., Choi, K. Y., Lee, J. J., Kim, B. C., Kwon, G. R., Lee, K. H., & Jung, H. Y. (2019). Ensembles of Patch-Based Classifiers for Diagnosis of Alzheimer Diseases. *IEEE Access*, *7*, 73373–73383. https://doi.org/10.1109/ACCESS.2019.2920011

Al-Zebari, A., & Sengur, A. (2019). Performance Comparison of Machine Learning Techniques on Diabetes Disease Detection. *1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019 - Proceedings*, 2–5. https://doi.org/10.1109/UBMYK48245.2019.8965542

Ayodele, T. O. (2012). Atherosclerotic Cardiovascular Disease. *Atherosclerotic Cardiovascular Disease*. https://doi.org/10.5772/711

Bagga, P., & Hans, R. (2015). Applications of mobile agents in healthcare domain: A literature survey. *International*

Journal of Grid and Distributed Computing, *8*(5), 55–72. https://doi.org/10.14257/ijgdc.2015.8.5.05

Bargarai, F. A. M., Abdulazeez, A. M., Tiryaki, V. M., & Zeebaree, D. Q. (2020). Management of wireless communication systems using artificial intelligence-based software defined radio. *International Journal of Interactive Mobile Technologies*, *14*(13), 107–133. https://doi.org/10.3991/ijim.v14i13.14211

Chitra, K. and. (2018). *Classification Of Diabetes Disease Using Support Vector Machine*. *3*(2), 1797–1801. https://www.researchgate.net/publication/320395340

Cinarer, G., & Emiroglu, B. G. (2019). Classificatin of Brain Tumors by Machine Learning Algorithms. *3rd International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2019 - Proceedings*. https://doi.org/10.1109/ISMSIT.2019.8932878

Conference, D. (1999). *Electric M ac h i nes and Drives Conference*.

Daniels, M., & Schroeder, S. A. (1977). Variation among physicians in use of laboratory tests II. Relation to clinical productivity and outcomes of care. *Medical Care*, *15*(6), 482–487. https://doi.org/10.1097/00005650-197706000-00004

Dey, A. (2016). Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies*, *7*(3), 1174–1179. www.ijcsit.com

Durai, V. (n.d.). *Liver disease prediction using machine learning*. *5*(2), 1584–1588.

Fadzilah Siraj, & Mansour Ali Abdoulha. (2011). Mining enrolment data using predictive and descriptive approaches. *Knowledge-Oriented Applications in Data Mining*, 53–72.

Fan, C. H., Hsu, Y., Yu, S. N., & Lin, J. W. (2013). Detection of myocardial ischemia episode using morphological features. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 7334–7337. https://doi.org/10.1109/EMBC.2013.6611252

Grimson, J., Stephens, G., Jung, B., Grimson, W., Berry, D., & Pardon, S. (2001). Sharing health-care records over the internet. *IEEE Internet Computing*, *5*(3), 49–58. https://doi.org/10.1109/4236.935177

Hariharan, M., Polat, K., & Sindhu, R. (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, *113*(3), 904–913. https://doi.org/10.1016/j.cmpb.2014.01.004

Hashem, S., Esmat, G., Elakel, W., Habashy, S., Raouf, S. A., ElHefnawi, M., Eladawy, M., & ElHefnawi, M. (2018). Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *15*(3), 861–868. https://doi.org/10.1109/TCBB.2017.2690848

Hazra, A., Kumar, S., & Gupta, A. (2016). Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms. *International Journal of Computer Applications*, *145*(2), 39–45. https://doi.org/10.5120/ijca2016910595

Hota, H. S. (2014). Identification of {Breast} {Cancer} {Using} {Ensemble} of {Support} {Vector} {Machine} and {Decision} {Tree} with {Reduced} {Feature} {Subset}. *Blue Eyes Intelligence Engineering & Sciences Publication Pvt. LtdInternational Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *3*(9), 4.

Hu, G., Qiu, Y., & Xiang, L. (2006). *Kernel-based reinforcement learning*. *4113 LNCS*, 757–766. https://doi.org/10.1007/978-1-4899-7687-1_100235

Huang, F. J., & LeCun, Y. (2006). Large-scale learning with SVM and convolutional nets for generic object categorization. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *1*(July 2006), 284–291. https://doi.org/10.1109/CVPR.2006.164

Huang, M. J., Chen, M. Y., & Lee, S. C. (2007). Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. *Expert Systems with Applications*, *32*(3), 856–867. https://doi.org/10.1016/j.eswa.2006.01.038

Iswanto, I., Laxmi Lydia, E., Shankar, K., Nguyen, P. T., Hashim, W., & Maseleno, A. (2019). Identifying diseases and diagnosis using machine learning. *International Journal of Engineering and Advanced Technology*, *8*(6 Special Issue 2), 978–981. https://doi.org/10.35940/ijeat.F1297.0886S219

Jahwar, A. F. (2021). *META-HEURISTIC ALGORITHMS FOR K-MEANS CLUSTERING : A REVIEW*. *17*(7), 1–20.

Javeed, A., Zhou, S., Yongjian, L., Qasim, I., Noor, A., & Nour, R. (2019). An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection. *IEEE Access*, *7*, 180235–180243. https://doi.org/10.1109/ACCESS.2019.2952107

Kharya, S. (2012). Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease. *International Journal of Computer Science, Engineering and Information Technology*, *2*(2), 55–66. https://doi.org/10.5121/ijcseit.2012.2206

Ko, J., Swetter, S. M., Blau, H. M., Esteva, A., Kuprel, B., Novoa, R. A., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. http://dx.doi.org/10.1038/nature21056

Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in Medicine*, *23*(1), 89–109. https://doi.org/10.1016/S0933-3657(01)00077-X

Kousarrizi, M. R. N., Seiti, F., & Teshnehlab, M. (2012). An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification. *International Journal of Electrical & Computer Sciences*, *12*(01), 13–19. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.655.363&rep=rep1&type=pdf

Kramer, O. (2013). Dimensionality Reduction with Unsupervised Nearest Neighbors. *Intelligent Systems Reference*

Library, *51*, 13–23. https://doi.org/10.1007/978-3-642-38652-7

Kulkarni, N. N., & Bairagi, V. K. (2017). Extracting Salient Features for EEG-based Diagnosis of Alzheimer's Disease Using Support Vector Machine Classifier. *IETE Journal of Research*, *63*(1), 11–22. https://doi.org/10.1080/03772063.2016.1241164

Kumar, D., Jain, N., Khurana, A., Mittal, S., Satapathy, S. C., Senkerik, R., & Hemanth, J. D. (2020). Automatic Detection of White Blood Cancer from Bone Marrow Microscopic Images Using Convolutional Neural Networks. *IEEE Access*, *8*(Mm), 142521–142531. https://doi.org/10.1109/ACCESS.2020.3012292

Liu, J., Xu, H., Chen, Q., Zhang, T., Sheng, W., Huang, Q., Song, J., Huang, D., Lan, L., Li, Y., Chen, W., & Yang, Y. (2019). Prediction of hematoma expansion in spontaneous intracerebral hemorrhage using support vector machine. *EBioMedicine*, *43*, 454–459. https://doi.org/10.1016/j.ebiom.2019.04.040

Mahmood, M. R., & Abdulazeez, A. M. (2019). Different Model for Hand Gesture Recognition with a Novel Line Feature Extraction. *2019 International Conference on Advanced Science and Engineering, ICOASE 2019*, *2018*, 52–57. https://doi.org/10.1109/ICOASE.2019.8723731

Mcculloch, W. S., & Pitts, W. (1990). A logical calculus nervous activity. *Bulletin of Mathematical Biology*, *52*(l), 99–115.

Meng, L., Ding, S., Zhang, N., & Zhang, J. (2018). Research of stacked denoising sparse autoencoder. *Neural Computing and Applications*, *30*(7), 2083–2100. https://doi.org/10.1007/s00521-016-2790-x

Mohammed, M., Khan, M. B., & Bashie, E. B. M. (2016). Machine learning: Algorithms and applications. In *Machine Learning: Algorithms and Applications* (Vol. 7, Issue 13). https://doi.org/10.1201/9781315371658

Murphy, K. P. (2012). Machine Learning - A Probabilistic Perspective - Table-of-Contents. *The MIT Press*, 1049.

Najim Adeen, I. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Systematic review of unsupervised genomic clustering algorithms techniques for high dimensional datasets. *Technology Reports of Kansai University*, *62*(3), 355–374.

Naqi, S. M., Sharif, M., & Jaffar, A. (2020). Lung nodule detection and classification based on geometric fit in parametric form and deep learning. *Neural Computing and Applications*, *32*(9), 4629–4647. https://doi.org/10.1007/s00521-018-3773-x

Padmapriya, B., & Velmurugan, T. (2015). A survey on breast cancer analysis using data mining techniques. *2014 IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC 2014*. https://doi.org/10.1109/ICCIC.2014.7238530

Pan, J., & Tompkins, W. J. (1985). Pan Tomkins 1985 - QRS detection.pdf. *IEEE Transactions on Biomedical Engineering*, *32*(3), 230–236.

Pölsterl, S., Conjeti, S., Navab, N., & Katouzian, A. (2016). Survival analysis for high-dimensional, heterogeneous medical data: Exploring feature extraction as an alternative to feature selection. *Artificial Intelligence in Medicine*, *72*, 1–11. https://doi.org/10.1016/j.artmed.2016.07.004

Rajabion, L., Shaltooki, A. A., Taghikhah, M., Ghasemi, A., & Badfar, A. (2019). Healthcare big data processing mechanisms: The role of cloud computing. *International Journal of Information Management*, *49*(May), 271–289. https://doi.org/10.1016/j.ijinfomgt.2019.05.017

Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B. W., Aslam, W., & Choi, G. S. (2020). COVID-19 Future Forecasting Using Supervised Machine Learning Models. *IEEE Access*, *8*, 101489–101499. https://doi.org/10.1109/ACCESS.2020.2997311

Vijayarani, S., & Dhayanand, S. (2015). Data Mining Classification Algorithms for Kidney Disease Prediction. *International Journal on Cybernetics & Informatics*, *4*(4), 13–25. https://doi.org/10.5121/ijci.2015.4402

Sadeeq, H., & Abdulazeez, A. M. (2018). Hardware Implementation of Firefly Optimization Algorithm Using FPGAS. *ICOASE 2018 - International Conference on Advanced Science and Engineering*, 30–35. https://doi.org/10.1109/ICOASE.2018.8548822

Sadiq, S. S., Abdulazeez, A. M., & Haron, H. (2020). Solving multi-objective master production scheduling model of Kalak refinery system using hybrid evolutionary imperialist competitive algorithm. *Journal of Computer Science*, *16*(2), 137–149. https://doi.org/10.3844/JCSSP.2020.137.149

Safavian, S. R., & Landgrebe, D. (1991). A Survey of Decision Tree Classifier Methodology. *IEEE Transactions on Systems, Man and Cybernetics*, *21*(3), 660–674. https://doi.org/10.1109/21.97458

Senturk, Z. K., & Kara, R. (2014). Breast Cancer Diagnosis Via Data Mining: Performance Analysis of Seven Different Algorithms. *Computer Science & Engineering: An International Journal*, *4*(1), 35–46. https://doi.org/10.5121/cseij.2014.4104

Shaheamlung, G., Kaur, H., & Kaur, M. (2020). A Survey on machine learning techniques for the diagnosis of liver disease. *Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020*, 337–341. https://doi.org/10.1109/ICIEM48762.2020.9160097

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888–905. https://doi.org/10.1109/34.868688

Singh, B. N., & Tiwari, A. K. (2006). Optimal selection of wavelet basis function applied to ECG signal denoising. *Digital Signal Processing: A Review Journal*, *16*(3), 275–287. https://doi.org/10.1016/j.dsp.2005.12.003

Smellie, W. S. A., Galloway, M. J., Chinn, D., & Gedling, P. (2002). Is clinical practice variability the major reason for differences in pathology requesting patterns in general practice? *Journal of Clinical Pathology*, *55*(4), 312–314. https://doi.org/10.1136/jcp.55.4.312

Stuart, P. J., Crooks, S., & Porton, M. (2002). An interventional program for diagnostic testing in the emergency department. *Medical Journal of Australia*, *177*(3), 131–134. https://doi.org/10.5694/j.1326-5377.2002.tb04697.x

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Transactions on Medical Imaging*, *35*(5), 1299–1312. https://doi.org/10.1109/TMI.2016.2535302

Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. T. (2020). Machine learning techniques for the diagnosis of alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications and Applications*, *16*(1s). https://doi.org/10.1145/3344998

V.S. Sriram, T., Rao, M. V., Narayana, G. V. S., & Kaladhar, D. S. V. G. . (2016). ParkDiag: A Tool to Predict Parkinson Disease using Data Mining Techniques from Voice Data. *International Journal of Engineering Trends and Technology*, *31*(3), 136–140. https://doi.org/10.14445/22315381/ijett-v31p223

Venkata Ramana, B., Babu, M. S. P., & Venkateswarlu, N. . (2011). A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis. *International Journal of Database Management Systems*, *3*(2), 101–114. https://doi.org/10.5121/ijdms.2011.3207

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning*, *July*, 1096–1103. https://doi.org/10.1145/1390156.1390294

Wennberg, J. E. (1984). Dealing with medical practice variations: A proposal for action. *Health Affairs*, *3*(2), 6–32. https://doi.org/10.1377/hlthaff.3.2.6

Wissel, B. D., Van Camp, P. J., Kouril, M., Weis, C., Glauser, T. A., White, P. S., Kohane, I. S., & Dexheimer, J. W. (2020). An interactive online dashboard for tracking COVID-19 in U.S. counties, cities, and states in real time. *Journal of the American Medical Informatics Association : JAMIA*, *27*(7), 1121–1125. https://doi.org/10.1093/jamia/ocaa071

Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. *IEEE Access*, *8*, 203097–203116. https://doi.org/10.1109/access.2020.3036072

Zebari, N. A., Zebari, D. A., Zeebaree, D. Q., & Saeed, J. N. (2021). Significant features for steganography techniques using deoxyribonucleic acid: a review. *Indonesian Journal of Electrical Engineering and Computer Science*, *21*(1), 338–347. https://doi.org/10.11591/ijeecs.v21.i1.pp338-347

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, *1*(2), 56–70. https://doi.org/10.38094/jastt1224

Zeebaree, D. Q., Abdulazeez, A. M., Zebari, D. A., Haron, H., & Hamed, H. N. A. (2020). Multi-level fusion in ultrasound for cancer detection based on uniform LBP features. *Computers, Materials and Continua*, *66*(3), 3363–3382. https://doi.org/10.32604/cmc.2021.013314

Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M. (2018). Gene Selection and Classification of Microarray Data Using Convolutional Neural Network. *ICOASE 2018 - International Conference on Advanced Science and Engineering*, *February*, 145–150. https://doi.org/10.1109/ICOASE.2018.8548836

Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019a). Machine learning and Region Growing for Breast Cancer Segmentation. *2019 International Conference on Advanced Science and Engineering, ICOASE 2019*, *April*, 88–93. https://doi.org/10.1109/ICOASE.2019.8723832

Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019b). Trainable Model Based on New Uniform LBP Feature to Identify the Risk of the Breast Cancer. *2019 International Conference on Advanced Science and Engineering, ICOASE 2019*, 106–111. https://doi.org/10.1109/ICOASE.2019.8723827

Zhuang, Z. Y., Churilov, L., Burstein, F., & Sikaris, K. (2009). Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. *European Journal of Operational Research*, *195*(3), 662–675. https://doi.org/10.1016/j.ejor.2007.11.003

Cite this article:

Nareen O. M. Salim & Adnan Mohsin Abdulazeez (2021). Human Diseases Detection Based On Machine Learning Algorithms: A Review. *International Journal of Science and Business*, *5*(2), 102-113. doi: https://doi.org/10.5281/zenodo.4467510

Retrieved from http://ijsab.com/wp-content/uploads/674.pdf

Published by

IJSAB International