MACHINE LEARNING AND DEEP LEARNING TECHNIQUES FOR CLASSIFICATION OF SKIN DISEASE DIAGNOSIS: A REVIEW

Article · January 2022

CITATION
1

READS
221

2 authors, including:

Kalaivani A.
NGM College
18 PUBLICATIONS **19** CITATIONS

SEE PROFILE

JOURNAL OF EDUCATION: RABINDRA BHARATI UNIVERSITY
ISSN : 0972-7175

MACHINE LEARNING AND DEEP LEARNING TECHNIQUES FOR CLASSIFICATION OF SKIN DISEASE DIAGNOSIS: A REVIEW

A. Kalaivani
Assistant Professor, Department of Computer Technology, Nallamuthu Gounder Mahalingam College , Pollachi &
Research Scholar, Department of Computer Science, PSGR Krishnammal College for Women., Coimbatore, Tamilnadu, India, kalaivanimathsca@gmail.com

Dr. S.Karpagavalli
Associate Professor, Department of Computer Science, PSGR Krishnammal College for Women Coimbatore, Tamilnadu, India, karpagavalli@psgrkcw.ac.in

ABSTRACT—
Skin disorders are caused by both genetics and environmental factors, they are the most frequent worldwide. Early detection of skin problems seems critical for successful treatment options. Detecting skin illness and determining its kind is a difficult challenge in the medical field. Because of the intricacy of human skin texture and the visual proximity impact of illnesses, determining the specific type might be challenging. As a result, it is critical to diagnose and recognize skin illness as quickly as possible. Artificial intelligence (AI) is quickly expanding in medical disciplines in today's world. Many machine learning (ML) and deep learning (DL) approaches are used for diagnostic reasons. These techniques significantly enhance and speed up the diagnosing procedure. This article provided a quick comparison of machine learning and deep learning approaches. Various image processing algorithms have been developed to identify skin cancer automatically as technology has progressed in recent years. Various strategies for categorizing skin lesions as benign or malignant are discussed in this study. We examined skin-related condition categorization using machine learning and image processing algorithms in this study.

Keywords— Deep learning, Image processing, Machine learning, Skin disease.

I. INTRODUCTION

There are many different types of skin problems, each with its own set of causes, such as internal ones involving hormones and body tissues, like acne, or external ones involving environmental damage or sun sensitivity, like blisters. Because of its complexity, diversity, and similarities, skin illness can only be diagnosed by dermatologists with substantial clinical experience, and it is seldom repeated. An untrained dermatologist is prone to misdiagnose it, which can worsen the problem and prevent proper treatment. Thus, a rapid and reliable approach to aid patients and dermatologists in data processing and evaluation is required. Skin diseases are frequent among many illnesses, and if these tactics are not fit for that form of skin condition, they will backfire. People are quickly infected by skin illnesses, which must be treated as early as possible.

Recent breakthroughs in computer vision using machine learning approaches have resulted in considerable improvements in computer - based diagnostic and detection systems for the early detection of severe malignant skin conditions. Deep learning developments have affected various scientific and industrial disciplines and resulted in important triumphs, drawing inspiration from the human nervous system. With deep learning's rapid growth in biomedical data processing, many professionals have turned

to it for more precise and trustworthy data. As the amount of available biomedical data, such as pictures, medical records, and omics, has grown fast, deep learning has seen substantial success in a range of medical image processing difficulties. The complexity of categorization is enhanced when each human skin characteristic is considered separately, and human extraction of characteristics is insufficient for classifying. Deep learning for disease detection has emerged as an exciting new research subject in dermatology. This study will cover how to use deep learning to examine the features of skin lesions and the current status of imaging technologies to detect skin illnesses. These deep learning-based algorithms, on the other hand, have a considerable training requirement, which is a huge number of annotated pictures for each class. Dermoscopy appears to be an image processing technology for reducing skin surface reflection and allowing in-person evaluation of problems inside the skin layers. With a better resolution, additional visual input from the skin layers may be obtained.

Machine learning is a field that can assist in predicting illness diagnosis using previously trained data. Various machine learning approaches have been developed by many scientists to identify various illnesses efficiently. Uses artificial intelligence systems to learn without the need for explicit programming. Machine learning techniques can be used to develop a model that can detect disease early and recommend treatments. Early detection and treatment are the most effective ways to reduce disease-related mortality rates. As a result, most clinical specialists are driven to innovative good predictive solutions for sickness prediction neural networks are computational.

Artificial neural networks understand through similarities in labeled data and then use inference to make meaningful predictions in the absence of precise recommendations. In the medical field, classification algorithms are widely used to better detect and forecast disorders. Diseases and health difficulties such as liver disease, development of complications, prostate cancer, diabetic, and cardiac dysfunction have quite a significant impact on someone else's health and can result in death if ignored. The healthcare business will make effective judgments by extracting hidden patterns and correlations from the database. Thanks to improvements in machine learning and artificial intelligence, various classification methods and clustering approaches, such as K-nearest, Logistic Regression, Stochastic Forests, SVM Classifiers, Nave Bayes, and others, can give a cure for the disease..

The most well-known of these deep learning model is CNN, which is designed to train image characteristics and detect skin disorders. Pre-trained CNNs would be used in an improved training model to explain the better pictures. In order to categories all sorts of input photos, such neural networks are trained on millions of photographs. Residual Network, AlexNet architecture, ImageNet, and VGGNet are examples of freely pre-learned neural network based models that may be used to execute various tasks.

The following is the overall structure of the article: The second section focuses into prior research on the classification of skin conditions. Section III provides an overview of machine learning methods and the Deep Convolutional Neural Networks Classifier, and Section IV shows the discussions. Section V includes the summary and offers recommendations for its future path.

II. LITERATURE SURVEY

Several researchers have proposed skin disease detection strategies based on imaging examinations. We will go through some of the techniques that have been mentioned in the literature at this point.

In 2018, Thiruna Vukkaras K. et al. [1] used a variety of different classifiers to diagnose liver illness.

Regression analysis, k-nearest neighbors, and multilayer perceptron were the approaches used in this study. The confusion matrix as well as the effectiveness score are used to compare these classification algorithms.

Melanoma cancer, for example, has such a higher risk of mortality but is treatable if caught early enough [2, 3]. Throughout the years, several computing technologies were also employed in the diagnosis and detection of tumors. A set of low-level pixel processing algorithms are often used in the traditional approach to medical picture analysis. In the pipeline process of melanoma cancer detection and diagnosis, primary processing methodologies such as images pre-processing, information extraction, extraction and classification, and categorization of tumors images were being detailed [4].

The Imagenet large scale visual recognition 2017 collection was employed in this study, which comprises around 2,000 increased dermatoscopic images divided into three basic categories: melanoma, nevus, and keratosis. Testing determines if these data are sufficient for accurately detecting and refusing normal and cancerous dysplasic diseases. Many epidermal histological analyzers have been suggested using Artificial Intelligence systems based on Deep Neural Networks (CNN), including the Imagenet, Resnet ensembles [5], R Convolutional neural networks [6], the Mask and Area Multi - Layer perceptron (Mask R CNN) and DeeplabV3+ method ensemble [7], and in overall numerous Convolutional constructions, creating this strategy one of the most powerful for effective feature investigation [8].

Inside the reviewed state-of-the-art, data from the Global Conference on Biomedical Imaging (ISBI) 2017 challenge is used as a reference as the most recent publication histopathologic object detection task (malignant/benign) for the ISIC (International Skin Imaging Collaboration) 2017 challenge. We undertake an objective comparison of data using the challenge's assessment criteria, such as precision, hypersensitivity, and accurate, using our recommended approach.

Therefore, we included a new statistic called balance accuracy, which was recommended for this particular instance of an imbalanced database [9-11]. A Receiver - operating characteristic Relation (ROC) vector is supplied to evaluate and display the efficiency of many different classifiers in a single map; this is a very useful statistic for evaluation in the healthcare profession, and its use in machine learning research has been increasing [12].

Researchers in [13] presented a detailed neural network for detecting skin lesions in input skin visuals using medical imaging. They developed a prototype device to diagnose skin diseases using approach. The goal of this research is to detect and diagnose skin disease using texture analysis based on thresholding and neural networks based on input skin images.

III. SKIN LESION IMAGE ANALYSIS TECHNIQUES

Computer vision (ML) is a machine intelligence subfield that allows robots to think and act such as people without the need for human interference [14]. It is kind of a method for making robots understand without having to be specifically designed. Machine Learning's major objective is to develop computer software that can acquire information and data from all this. Machine learning may take numerous forms.

Each sort of machine learning will be briefly defined in this section. Using supervised learning, the algorithm learns by adjusting input data variables, which includes both data and information. The method is given simple input data and told to learn the underlying structure from it in unsupervised learning [15]. Furthermore, both strategies are required for semi-supervised learning (various classification

environments). They employ both labelled and extracted features, which implies that some input has specific information needs while some do not. In reinforcement learning, the system tries to learn from its surroundings by rewarding desired actions and punishing undesirable ones [16]. The usage of these technologies is employed for a variety of medical activities, including illness diagnosis [17]. Artificial intelligence is a branch of machine learning that consists of several levels that hold different levels of observation, with every layer receiving number of inputs and transferring the result to another.

A number of machine learning methods are used to identify the disease. The following are the majority of them that were detected in this review paper:

3.1 K - Neighbourhood Algorithm (KNN)

KNN is a fundamental training algorithm that is often used for categorization, pattern recognition, and prediction. The Distance measure between datasets is used by KNN to discover data neighbours. This approach is then used to address classification and regression problems. Furthermore, the k value (where k is an application variable) will detect all initial feature cases that are suitable for the new case and will surround all occurrences in order to obtain the incident for a similar place [18]. As a result, the Value of k is important and should be carefully planned, as a low K value could result in over-fitting. There are certain drawbacks, such as low performance when dealing with a large training dataset. Because we must calculate the distance between all training samples and each query instance, the computation cost is quite large.

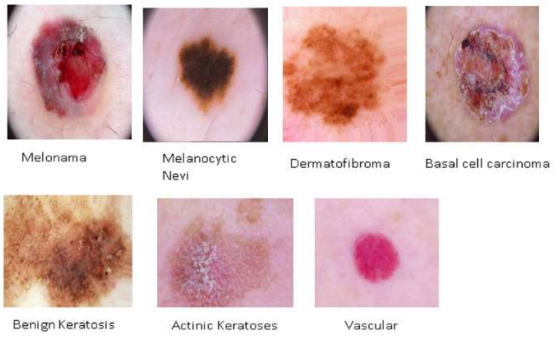

Figure 1. Example dermoscopy images for skin sores of HAM Dataset

3.2 Support Vector Machine

The Classification Method has been shown to be useful for a variety of classification applications. [19]. The number of points on the edge of the class descriptors is used to find the best hyperplane between classes. The gap separating two classes is known as the margins. The categorization accuracy increases with a bigger margin. The data sets on the border are called support vectors. Both regression and classification issues are solved using SVM. Furthermore, shrinking the image shortens the processing time and hence improves system performance.

Classifier divides pieces of data into two groups by searching for features in a high-dimensional space and picking the best hyper-plane. It handles both small and large data effectively.

3.3 Naive Bayes algorithm

Naive Bayes (NB) is a statistical and statistical method based on a classification algorithm. It is a common strategy in neural network models because of its simplicity in enabling all features to contribute equally to the final choice. The NB technique is intriguing and applicable for a variety of applications because of its computational efficiency. Prior, posterior, and class conditional probability are crucial components of the NB classification [20]. This approach has several advantages, including being simple to apply with large datasets and being quite effective. It might be used to solve binary and multiclass classification issues. A lesser quantity of training data is required, and both discrete and continuous data can be employed. This computation application is also used to filter spam emails and categories studies [21].

3.4 Algorithm of Decision Trees

The Decision Tree (DT) is a supervised machine learning approach for solving regression and classification tasks that splits data periodically based on a parameter.

The data is grouped into clusters, with the final conclusions represented by the tree's leaf. The purpose of the logistic regression is to employ training information to analyze fundamental choice rules in order to develop a model for predicting the predicted value [22]. The tree is formed using the data for training during the training stage. The name of the class is stored in the leaf nodes, but not in the decision nodes. The decision tree is used to arrange quantitative and qualitative criteria. The nonlinear link between arguments has no effect on the tree's efficiency.

3.4 Deep Learning Classification Technique

Nonlinear transformations are used in transfer learning, which a subset of machine is learning. It uses a variety of algorithms to understand how to comprehend incoming data using several levels of processing and complicated structures. Deep learning designs include pre - trained model (Recurrent neural network), deep auto - encoders, and artificial neural network (ANN). Speech recognition, image recognition, and pharmaceutical industries are among the sectors where these techniques have been applied. Recurrent neural networks is a neural training process that stores recent data in internal memory. In RNN architecture, memory units in general contain linkages to each other that transmit information from the previous execution. RNN modifies the present forward process in order to adapt to the context of current input [23]. Because the region of interest in MRI medical pictures is often dispersed across multiple neighbouring slices, subsequent slices have commonalities. As a kind of sequential data, from the input slices, RNN may extract inter-slice contexts [24].

Other computational intelligence technique, the convolutional neural network (CNN), is used to solve complex problems and categorize large amounts of data. The model is made up of four layers: convolutional, max-pooling, completely connected, and the output layer. CNN structure for medical image classification with multi classes that accepts a 32-bit sub picture from the medical image Pooling layers minimize the number of feature maps to be sent to the succeeding layers, allowing the fully connected layer to deliver information as a predictions of the appropriate category [25].

Table 1. The Advantage and Disadvantage of the Mentioned Machine Learning Techniques in Section Three

Machine - Learning Techniques	Merits	Demerits
KNN	A model is inexpensive and simple to apply. Both classification and regression are possible with this method. On multiclass issues, it runs nicely.	The amount of computation is enormous. The cost of classifying unknown records is high. High sensitivity to unnecessary information.
SVM	It can handle both linear and nonlinear data sets. Over-fitting is less likely to occur. Big data may be scaled up.	When working with a huge dataset, performance suffers. Selecting an appropriate kernel function is difficult. When a dataset is noisy, they don't operate well.
Naive Bayes	The simple option for large datasets. Deals with both discrete and continuous data. Both binary and multi-classification are possible. Irrelevant characteristics have no effect.	Models with several variables are very computationally expensive. Models that have been properly trained and tweaked can sometimes outshine Nave Bayes models, which are too simplistic.
Decision Tree	Both regression and classification are possible with this method. Handling numeric and categorical data is simple.	Over-fitting may arise as a result of the tree's repetitive building. Larger trees are more difficult to read.
Deep Learning	Features are automatically detected. It may be used with a variety of data types.	GPUs are required for training. Because of the complicated data models, training is quite expensive.

IV. DISCUSSION

All of the publications discussed in the preceding section are related research from 2009 to 2021 that employed various machine learning methods to diagnose illnesses. The evaluated work uses a variety of approaches, including K-nearest neighborhood, SVM classifier, logistic regression, and ensemble model. These methods were also applied to standard datasets in a number of disorders.

In [26] and [27], the K - means and SVM classifiers were used to different datasets and attributes. In terms of performance, the SVM approach beat the KNN algorithm [26]. In contrast, the KNN method outperforms SVM by roughly 4% in [27]. The prediction tree's efficiency is around 1% greater than the uniform distribution, according to study [27]. The accuracy of the ensemble model utilized by researchers in [28] and [29] was 98.50 percent and 90 percent, respectively. SVM is the sole binary classifiers, but decision trees, KNN, nave Bayes, and CNN can classify into more than two groups.

When the dataset is small, all algorithms function well, however deep learning techniques such as CNN

are superior for huge datasets. To recap, the technique's accuracy is determined by the size of the dataset and the classification algorithm, and the network method outperforms the statistical model.

V. CONCLUSION

In the medical industry, algorithms has become popular for providing tools and analyzing sickness data. As a result, learning algorithms are crucial for early disease detection. This study looked at a variety of machine learning algorithms for sickness prediction, as well as standard datasets, in illnesses such skin disease, breast cancer, heart disease, skin malignancies, and many more. Several methods for predicting SVM, K-nearest neighbours, and the logistic regression were determined to have high precision after evaluating several studies for various disease prediction models. However, because various important aspects, such as data, extraction of features, and the number of attributes, impact model accuracy and performance, the efficiency of the same technique may differ from one database to the next. Another noteworthy discovery in this review is that the accuracy and effectiveness of an ensemble classifier may be enhanced by using a different approach to generate it.

ACKNOWLEDGMENTS

I convey my sincere thanks to the management of Nallamuthu Gounder Mahalingam College, Pollachi, for allowing me to carry out this research work by funding with SEED MONEY. I extend my sincere thanks to the Principal, Deans, HoD and Manager of Nallamuthu Gounder Mahalingam College, Pollachi for providing the necessary facilities to carry out the present work.

REFERENCES

[1] Thirunavukkarasu K, Ajay S. Singh, Md Irfan and Abhishek Chowdhury, "Prediction of Liver Disease using Classification Algorithms", International Conference on Computing Communication and Automation (I.C.C.C.A.), IEEE, 2018.

[2] Al-Masni MA, Al-antari MA, Choi M-T, Han S-M, Kim T-S (2018) Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Comput Methods Programs Biomed 162:221–231.

[3] Balch CM, Gershenwald JE, Soong S, Thompson JF, Atkins MB, Byrd DR, Buzaid AC et al (2009) Final version of 2009 AJCC melanoma staging and classifcation. J Clin Oncol 27(36):6199.

[4] Unver HM, Ayan E (2019) Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. Diagnostics 9(3):72.

[5] Harangi B. Skin lesion classifcation with ensembles of deep convolu-tional neural networks. J Biomed Inform. 2018;86:25–32. https://doi.org/10.1016/j.jbi.2018.08.006.

[6] Nida N, Irtaza A, Javed A, Yousaf MH, Mahmood MT. Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. Int J Med Inform. 2019. https://doi.org/10.1016/j.ijmedinf.2019.01.005.

[7] Goyal M, Yap MH. Automatic lesion boundary segmentation in dermo-scopic images with ensemble deep learning methods. 2019. http://arxiv.org/abs/1902.00809. Accessed 16 July 2019.

[8] Jafari MH, Nasr-Esfahani E, Karimi N, Soroushmehr SMR, Samavi S, Najarian K. Extraction of skin lesions from non-dermoscopic images using deep learning. Int J Comput Assist Radiol Surg. 2016. https://doi.org/10.1007/s11548-017-1567-8.

[9] Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. In: 20th International conference on pattern recognition. 2010; p. 3121–24. https://doi.org/10.1109/ICPR.2010.764.

[10] R. J. Urbanowicz, J. H. Moore (2015) ExSTraCS 2.0: description and evaluation of a scalable learning classifer system. Evol Intell 8(2–3): 89–116. DOI: https://doi.org/10.1007/s12065-015-0128-8.

[11] Velez DR, et al. A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet Epidemiol. 2007;31(4):306–15. https://doi.org/10.1002/gepi.20211.

[12] Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett. 2006. https://doi.org/10.1016/j.patrec.2005.10.010.

[13] Ambad, Pravin S., and A. S. Shirat, "A Image Analysis System to Detect Skin Diseases," IOSR Journal of VLSI and Signal Processing (IOSR-JVSP), Volume 6, Issue 5, Ver. I (Sep. - Oct. 2016), PP 17-25.

[14] Golmei Shaheamlung, Harshpreet Kaur and Mandeep Kaur, "Survey on machine learning techniques for the diagnosis of liver disease", International Conference on Intelligent Engineering and Management, IEEE, 2020.

[15] AJAY SHRESTHA and AUSIF MAHMOOD, "Review of Deep Learning Algorithms and Architectures", IEEE, 2019.

[16] Yuanyuan Jia, Zhiren Tan and Junxing Zhang, "DKDR: An Approach of Knowledge Graph and Deep Reinforcement Learning for Disease Diagnosis", IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking, 2019.

[17] Meherwar Fatima and Maruf Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic", Journal of Intelligent Learning Systems and Applications, 2017.

[18] Animesh Urgiriye and Rupali Bhartiya," Review of Machine Learning Algorithm on Cancer Data Set", International Journal of Scientific Research & Engineering Trends, Volume 6, Issue 6, Nov-Dec-2020.

[19] Adnan Mohsin Abdulazeez, Maryam Ameen Sulaiman, Diyar Qader Zeebaree, "Evaluating Data Mining Classification Methods Performance in Internet of Things Applications", Journal Of Soft Computing And Data Mining, Vol.1, NO. 2, pp. 11-25, 2020.

[20] Aji Prasetya Wibawa and et al, "Naive Bayes Classifier for Journal Quartile Classification", IJES – Vol. 7, No. 2, 2019.

[21] Shahadat Uddin, Arif Khan, Md Ekramul Hossain and Mohammad Ali Moni3, "Comparing different supervised machine learning algorithms for disease prediction", B.M.C. Medical Informatics and Decision Making, 2019.

[22] Yash Jayesh Chauhan, "Cardiovascular Disease Prediction using Classification Algorithms of Machine Learning", International Journal of Science and Research (I.J.S.R.), ISSN: 2319-7064, 2018.

[23] Tarek M. Hassan1, Mohammed Elmogy and El-Sayed Sallam, "Diagnosis of Focal Liver Diseases Based on Deep Learning Technique for Ultrasound Images", pulished onlinee, springer, 2017.

[24] Siwei Ma, Xinfeng Zhang, Chuanmin Jia, Zhenghui Zhao, Shiqi Wang, and Shanshe Wang, " Image and Video Compression with Neural Networks: A Review", IEEE Transactions On Circuits And Systems For Video Technology, 2019.
[25] Mohammad Hesam Hesamian,Wenjing Jia, Xiangjian He and Paul Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges", Journal of Digital Imaging, Springer, 2019.
[26] Ankita Tyagi, Ritika Mehra and Aditya Saxena, "Interactive Thyroid Disease Prediction System Using Machine Learning Technique", 5th IEEE International Conference on Parallel, Distributed and Grid Computing(PDGC-2018), Solan, India, 20-22 Dec 2018.
[27] Noor Basha, Ashok Kumar P S, Gopal Krishna C and Venkatesh P, "Early Detection of Heart Syndrome Using Machine Learning Technique", 4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), 2019.
[28] Naresh Khuriwal and Nidhi Mishra, "Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm", IEEE, 2018.
[29] Rahma Atallah and Amjed Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method", IEEE, 2019.