ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2023.0405

SENTIMENT ANALYSIS IN MELANOMA CANCER DETECTION USING ENSEMBLE LEARNING MODEL

M. Uma Maheswari and A. Aloysius
Department of Computer Science, St. Joseph's College, Affiliated to Bharathidasan University, Tiruchirappalli, India

Abstract
Machine learning has the potential to improve healthcare by allowing clinicians to spend more time caring for patients and less time diagnosing them. This would allow clinicians to spend more time improving patient quality of life. Consequently, it is able to compute the risk of melanoma on a patient level and advise users to schedule a medical checkup rather than evaluating whether or not a specific lesion image that is provided by a patient is malignant. This is because the result of this is that it is able to compute the risk of melanoma at the patient level. By doing so, both the credibility and legislation issues are resolved, and the application is transformed into one that is adaptable. In this paper, we develop a machine learning ensemble to classify the melanoma cancer. The simulation is conducted in terms of training, testing accuracy, precision and recall. The results show that the proposed method achieves higher classification rate than other methods.

Keywords:
Machine Learning, Ensemble, Prediction, Melanoma

1. INTRODUCTION

Melanoma is one of the forms of cancer that affects individuals in Australia at an alarmingly high rate [1]. Despite the fact that it is fatal, there is a nearly 100 percent chance of recovery if treatment is started early enough. Therefore, its early discovery is essential to the process of saving the lives of patients. Even though dermatologists now have the ability to establish a diagnosis based on biopsies that is quite accurate, the technique is still invasive and painful [2]. However, because of the many different shapes that lesions might take, it is not possible to guarantee a high degree of accuracy [2].

Researchers, on the other hand, have used AI technologies such as machine learning (ML) and, more specifically, deep learning (DL) approaches to achieve exceptional accuracy in melanoma detection and classification [3]. Artificial intelligence (AI) technologies have not yet found broad usage in dermatology, even though they are far more effective and produce no negative side effects.

To begin, even though ML and DL classifiers are still considered to be relatively new technologies, a significant number of researchers have done experiments and authored publications on the topic.

However, a significant number of them were published without having been evaluated by any other authorities [4], which calls into question the dependability of these approaches. In the world of medicine, there is little question that having a false impression of high accuracy might have catastrophic consequences. The legislative bottleneck is another issue that won't be resolved any time soon. If representatives want to see machine learning techniques used in hospitals, they will need to find a way to strike a balance between the benefits the technology offers and the hazards it could bring [4].

As a result, the use of machine learning for melanoma detection will be discontinued by clinicians in the not-too-distant future. Even if there are several concerns that need to be addressed before ML and DL can be universally implemented as professional medical techniques, there are still realistic prospects for their use in the field of personalized healthcare. These prospects include machine learning and deep learning.

2. RELATED WORKS

A crawler was used to collect individual status updates from Facebook. A crawler that gathers user Facebook status updates was powered by tokens that Facebook Graph APQ users donated to the platform [5].

The challenge of predicting depression in the study can be stated as a binary classification problem, with the existence or absence of a depression diagnosis serving as labels. This was found in their research. Researchers have employed a features-based technique for gathering the data that can be used to determine links between the author writing style and emotional distress to overcome the problem that was caused by machine learning. This has allowed them to solve the problem [6].

A multi-modal depressive dictionary learning strategy was offered as a method for identifying depressed users of Twitter. The difference in online behavior between depressed and non-depressed users on social media was revealed by analyzing the contribution of the feature modalities and detecting depressed users on a large-scale depression candidate dataset. This allowed for the discovery of differences in online behavior between depressed and non-depressed users on social media [7].

According to the findings of a meta-regression analysis that are carried out, the overall sensitivity and the probabilities of identification in a univariate model are significantly affected by the method of documentation of the regression, the age of the sample, and the date on which the study was published. Once the age of the samples and the publishing date were factored into the model, it was discovered that the documentation technique had a significant impact on the overall sensitivity. When it comes to accurately diagnosing depression, primary care physicians and other types of specialists have a dismal track record [8].

The Knowledge-Based Recommendation System (KBRS), which includes an emotional health monitoring system to identify individuals who may be experiencing psychological issues such as stress or depression. The ontology- and sentiment-based KBRS will then target users who are having psychological challenges and send them encouraging, reassuring, or inspiring messages. In addition, the system is equipped with a notification system that will send an alarm to the appropriate authorities if it identifies a disruption associated with depression [9].

It is possible that the methodology proposed will be suitable for studies analyzing the efficacy of various training programs for physicians with the goal of increasing the precision with which they diagnose depression. The thoughts that people express on the internet might shed insight on the behaviors and routines that they engage in. The ability to recognize extremist content is a huge help when trying to Fig.out how users feel about a certain extremist group and how to prevent them from engaging in criminal action. It also makes it much easier to know how to discourage users from engaging in illegal behavior [10].

3. PROPOSED METHOD

In order to evaluate how successful, the proposed model would be, three health and medical datasets sourced from Twitter and UCI will be utilized.

3.1 DATA PRE-PROCESSING

The real-world data that has been collected from the various data communities has a significant amount of background noise, this data needs to be pre-processed in order to extract the important features that are necessary for modeling. Since most corpora are still in their original, unprocessed state, it is very necessary to first clean the data. First, we removed every case that featured an emoticon or a URL, and then we analyzed each one to determine its significance. The hash tags have been replaced with their respective definitions.

The processing of jargon and abbreviations that are frequently misspelled is made possible by the utilization of regular expressions. At this point in the processing, we get rid of all the terms that aren't very informative and that would otherwise make our corpus look cluttered. Sanitization, the elimination of stop words, and tokenization are some of the pre-processing techniques that have been utilized to cut down on the amount of data contained in the corpus.

During the process of sanitization, the text is converted from upper to lower case so that any remnants of a numerical value are concealed. To strip the words of their inflected endings, a method referred to as lemmatization was utilized. The process of removing stop words from written English involves removing all English words that do not provide any new or helpful information. We developed a third set of wordlists that included 53 unique stop words along with an extra 1200 extremely positive words and 1800 extremely negative terms.

This was done to supplement the positive, negative, and stop-word lists that were already in existence. These lexicons were developed so that the influence that the various emotional concepts had on the data could be better comprehended. The words contained in the word lists are then manually annotated with emoticons and labels denoting positive, extremely positive, negative, and extremely negative sentiments.

Because of the emotional weight that is associated with emoticons, it is common practice for people to manually insert them into tweets. As a direct consequence of this, a lexicon of 140 emoticons was developed, and each one was assigned to one of four distinct categories: positive, negative, highly positive, and extremely negative.

3.2 FEATURE EXTRACTION AND SELECTION

To extract the attributes, N-gram tokenization, which involves breaking the text into a series of tokens, was performed after pre-processing. This was done to accomplish the goal. N-grams are a group of tokens that appear repeatedly in a frame and are often used for token prediction. N-grams can be found by looking at frames. The process of extracting opinion terms from the N-gram lexicon is advanced by doing annotation and summarization at the sentence level. This N-gram dictionary is manually annotated using the pre-processed data, and during the process of summarizing, filler words such as a, and the, there, etc. were deleted from the review sentences. This was done to ensure that no important information was overlooked. The frequency distribution of sentence summaries is displayed in Fig.4.

Every word, including emoticons, was assigned a score; however, the scores for positive and negative keywords were not the same. In a subsequent step, the opinion terms were pulled out, feature vectorization was performed on them, and then they were transformed into vectors. The Term Frequency and Inverse Document Frequency (TF-IDF) measure was used to assign a score to each feature vector in the dataset. The equation is utilized to determine the relative weight of each feature vector.

$$TFIDF(t,d,D)=TF(t,d)\cdot IDF(t,D)$$

$$IDF(t,D)=\log N|\{d\in D:t\in d\}|$$

where, t – occurrence of term in a document d. N - total documents available in corpus and $|\{d\in D:t\in d\}|$ - term 't' appears in the total number of documents 'D'.

Based on the outcome of this calculation, the pipeline will proceed to model the weighted feature vectors by employing the Latent Dirichlet Allocation technique.

3.3 TOPIC MODELLING

An unsupervised probabilistic method is utilized in topic modeling using latent dirichlet allocation (LDA), which is used to model a corpus into a collection of themes. After linear discriminant analysis (LDA) is carried out on the feature vectors, a hierarchical Bayesian technique is utilized in this work to locate the text that is contained inside the corpus.

Let imagine that a corpus that is strong in reviews also has t-topic groups. Each review in the corpus is arbitrarily created by k subjects, and the topics themselves can be thought of as a polynomial probability distribution of feature vectors. Additionally, the reviews are organized into a hierarchical structure. The process of feature extraction when using LDA can be summarized that can be found below.

$$P(f_i) = \sum_{k=1}^{N} P\left(f_i t_i = k\right) P\left(t_i = k\right)$$

where, $P(t_i=k)$ - k topic probability, $P(f_i|t_i=k)$ - f_i probability of under k and N - total topics.

A feature vector known as f is produced because of merging the topics denoted by t with those denoted by r in corpus c and drawing samples from the word distributions.

If we assume that $\phi_k=P(f_i|t_i=k)$, where P is a multinomial distribution of feature vectors for topic k, the previous equation can be simplified even further. This will allow us to solve the problem more easily. The notation $r=P$ is used to indicate a multinomial distribution of review topics t. It is estimated that the parameters and are utilized when conducting reviews as well as feature vector analyses.

Here, r stands for review and Nr indicates the total number of attributes. In this instance, R refers to the overall number of assessments. Both the topic-word and review-topic Dirichlet distribution hyperparameters, as well as the review-topic Dirichlet distribution, are required for the formation of a corpus. r is a review-level variable that is sampled just once for each review, in contrast to, which is a feature-level variable that is sampled once for each word in r with Nr.

However, it is difficult to provide an express score for any of the properties involved. The Gibbs sampling method is utilized to circumvent this restriction and obtain the required parameter values. It then creates distributions of topic probabilities by counting the number of subjects that are included in the review corpus and using the data that is gathered from reviews and features. The following is an example of the conditional probability that is used by the Gibbs sampler.

LDA is utilized in this proposed research project to search through the review corpus texts and integrate them into latent topics. We have simulated scenarios on the review corpus, with K equal to 100, 200, 300, 400, and 500 different subject areas. The most important elements were gleaned from the data by making use of the probabilities connected to the various topics and characteristics. The next part will examine the use of feature selection as a strategy for overcoming the problems caused by the existence of an infinite number of dimensions.

3.4 FEATURE SELECTION

To discover the characteristics that will be incorporated into the final model, the LDA topic modeling is mined using a chi-square selection in conjunction with a principal component analysis. As a result of the huge number of dimensions (topics) that are produced by LDA, a feature selection model is necessary in order to combat the curse of dimensionality. Both approaches reduce the number of available features, which can speed up the learning process and improve disposition.

The chi-square selection is used to determine which characteristics are the most essential across all dimensions. Next, a principal component analysis is conducted to reduce the focus of the investigation. Principle components, which are variables that are not correlated with one another, are acquired through the application of principal component analysis (PCA) to calculate orthogonal transformations of the variables that define the dimensions of the existing features. Because principal component analysis places its emphasis on variability rather than correlation, the chi-square test is the method of choice for locating strongly correlated variables.

3.5 ENSEMBLE CLASSIFICATION

The fundamental idea behind an ensemble classification model can be broken down into two stages, as illustrated in Fig.2: the first stage involves obtaining classification results by using many weak classifiers, and the second stage involves integrating multiple classification results into a consistency function in order to obtain the final result by voting methods. Examples of well-known ensemble classification strategies include bagging, AdaBoost, random forest, random subspace, and gradient boosting. The bagging method provides sample subsets for training the basic models for integration by randomly sampling from the training data set. This allows the bagging method to train the basic models. During the training of the core models that comprise the Bagging model, parallelism is applied.

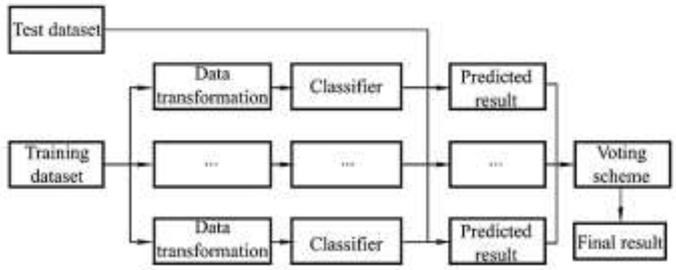

Fig.1. Ensemble Classification

4. RESULTS AND DISCUSSIONS

Several studies have been carried out on our end in order to determine whether or not the ensemble model is superior to the other options. In addition, we demonstrate that the reliability of the sources from which medical photographs originate, in addition to the quality of the images themselves and their accessibility, are significant components in the process of cancer detection. In the course of this research, one of the datasets that was used was called Melanoma TFRecords 256x256 (https://www.kaggle.com /datasets/cdeotte/melanoma-256x256). These TFRecords contain both an image from the melanoma classification competition as well as data in tabular format.

Table.1. Training and Testing Results

Methods	Training	Testing
Logistic Regression	39.12	63.43
C-Support Vector	42.13	64.96
Support Vector Regression	51.15	77.24
KNN Classifier	60.17	78.26
Decision Tree	60.17	79.28
Random Forest	63.18	80.82
AdaBoost	66.20	80.82
Gradient Boosting	66.20	82.35
Bernoulli Naïve Bayes	66.79	83.37
Gaussian Naïve Bayes	70.40	89.00
Linear Discriminant Analysis	72.21	90.02
Proposed Classification	92.21	94.12

Table.2. Precision and Classification Error

Methods	Precision	Error
Logistic Regression	61.99	61.99

C-Support Vector	62.66	69.49
Support Vector Regression	79.91	79.91
KNN Classifier	80.61	80.61
Decision Tree	86.03	86.03
Random Forest	87.21	87.21
AdaBoost	92.40	92.40
Gradient Boosting	92.87	92.63
Bernoulli Naïve Bayes	92.87	92.87
Gaussian Naïve Bayes	93.35	92.87
Linear Discriminant Analysis	97.35	97.35
Proposed Classification	99.00	99.00

Identification of cancer at an early stage is necessary in order to lower mortality rates caused by cancer. Currently, medical imaging, blood tests, fine-needle aspiration cytology (FNAC), and excisional biopsies are used to manually identify thyroid disease; however, this method is inefficient, time-consuming, and may be impacted by unreliable human false-positive rates. The development of ensemble learning has been of tremendous assistance in the field of disease diagnosis. The research that is being done currently on the classification of ultrasound images is inefficient since it requires medical professionals to manually designate the region of interest, then do feature extraction and selection, and then classify individual nodules.

5. CONCLUSIONS

In this paper, the degrees of precision range widely, from 42% to 99% and the accuracy tends to get better as the size of the dataset gets larger, particularly when employing ensembles. Ensemble learning after the extracted feature as input and output classification, have also been demonstrated to be reasonably good in the categorization of melanoma. In addition, because of the development of deep learning, cutting-edge models that can be used in the diagnosis of thyroid cancer have been created and validated in the field. Considering this, we have developed a flexible framework that is comprised of ensemble learning. This framework has been put to the test on real-world data sets and has been found to perform better than earlier research in the field of cancer diagnosis. In addition, the data can be simply interpreted, which, because of the outputs that were chosen, makes it possible for clinical applications to be implemented.

REFERENCES

[1] V.R. Allugunti, "A Machine Learning Model for Skin Disease Classification using Convolution Neural Network", *International Journal of Computing, Programming and Database Management*, Vol. 3, No. 1, pp. 141-147, 2022.

[2] Leilei Sun, Chuanren Liu, Chonghui Guo, Hui Xiong and Yanming Xie, "Data-driven Automatic Treatment Regimen Development and Recommendation", *Proceedings of International Conference on Knowledge Discovery and Data Mining*, pp. 1865-1874, 2016.

[3] R. Manikandan and M. Ramkumar, "Sequential Pattern Mining on Chemical Bonding Database in the Bioinformatics Field", *Proceedings of International Conference on AIP*, Vol. 2393, No. 1, pp. 1-13, 2022.

[4] C. Alvino Rock and R.J.S. Jeba Kumar, "Computer Aided Skin Disease (CASD) Classification using Machine Learning Techniques for iOS Platform", *Tracking and Preventing Diseases with Artificial Intelligence*, pp. 201-216, 2022.

[5] B.R. Nanditha, "Oral Cancer Detection using Machine Learning and Deep Learning Techniques", *International Journal of Current Research and Review*, Vol. 14, No. 1, pp. 64-78, 2022.

[6] N. Sultana, "Predicting Sun Protection measures against Skin Diseases using Machine Learning Approaches", *Journal of Cosmetic Dermatology*, Vol. 21, No. 2, pp. 758-769, 2022.

[7] Y.N. Chen and W.B. Wei, "Machine Learning Models for Outcome Prediction of Chinese Uveal Melanoma Patients: A 15-Year Follow-Up Study", *Cancer Communications*, Vol. 42, No. 3, pp. 273-276, 2022.

[8] A.R. Khan, "Facial Emotion Recognition using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges", *Information*, Vol. 13, No. 6, pp. 268-278, 2022.

[9] M. Pinto, A. Ammendolia and A. De Sire, "Quality of Life Predictors in Patients with Melanoma: A Machine Learning Approach", *Frontiers in Oncology*, Vol. 12, pp. 843611-843618, 2022.

[10] N. Sengupta and U. Ghone, "Scarcity of Publicly Available Oral Cancer Image Datasets for Machine Learning Research", *Oral Oncology*, Vol. 126, pp. 105737-105743, 2022.