International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 6, October 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Skin Lesion Classification using Supervised Algorithm in Data Mining

G. Saranya[1], Dr. S. M. Uma[2]

[1]M.E (CSE) Student, [2]M.E., (Ph.D), Assistant Professor,
[1,2]Department of CSE, Kings College of Engineering, Thanjavur, Tamil Nadu, India

ABSTRACT

Skin cancer is one of the major types of cancers with an increasing incidence over the past decades. Accurately diagnosing skin lesions to discriminate between benign and skin lesions is crucial.J48 Algorithm and SVM (SUPPORT VECTOR MACHINE) based techniques to estimate effort. In this work proposed system of the project is using data mining techniques for collecting the datasets for skin cancer. So that system can overcome to diagnosing the disease quickly and accuracy. Comparing to other algorithm proposed algorithm has more accuracy. When we have to using two kind of algorithm .They are J48, SVM. J48 Algorithm produced better accuracy more than SVM algorithm. The accuracy of the proposed system is 90.2381%. It means this prediction is very close to the actual values.

KEYWORDS: *Support Vector Machine, J48 Algorithm, SVM algorithm, data mining techniques*

How to cite this paper: G. Saranya | Dr. S. M. Uma "Skin Lesion Classification using Supervised Algorithm in Data Mining" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-6, October 2019, pp.1146-1148, URL: https://www.ijtsrd.com/papers/ijtsrd29346.pdf

IJTSRD29346

INTRODUCTION

PROBLEM: Cancer disease is a major type of disease among people worldwide. Many People affected bycancer. Often kind of people affected by skin lesion disease Even though they are using more protectingcreams. The early stage skin lesion cancer is not easily identifiable, which is the main reason for thisskin cancer is people used in more and more unknowing cream without doctor prescription .so that iswhy the people knowingly or unknowingly affected by skin lesion cancer.

SOLUTION: To overcome this problem we design an efficient, we are going to apply the classificationby several algorithm in skin lesion disease classified using supervised algorithm . They are J48 andSVM algorithm In this system, we have to finding ,how much possibilities are there to curingskin lesion disease in early stage of cancer.

Weka is open source software for data mining under the GNU General public license. This system is developed at the University of Waikato in New Zealand. "Weka" stands for the Waikato Environment for knowledge analysis. Weka is freely available at http://www.cs.waikato.ac.nz/ml/weka. The system is written using object oriented language java. Weka provides implementation of state-of-the-art data mining and machine learning algorithm. User can perform association, filtering, classification, clustering, visualization, regression etc. by using weka tool.

Each and every organization is accession vast and amplifying amounts of data in different formats and different databases at different platforms. This data provides any meaningful information that can be used to know anything about any object. Information is nothing just data with some meaning or processed data. Information is than converted to knowledge to use with KDD. Data Mining is a non trivial extraction of implicit, previously unknown, and imaginable useful information from data. Data mining finds important information hidden in large volumes of data.

Data mining is the reasoning of data. It is the use of software techniques for finding patterns and consistency in sets of data. Data Mining is an interdisciplinary field involving: Databases, Statistics, and Machine Learning. There are various techniques available for data mining as given below:-

A. **Association Rule Learning: -** This is also called market basket analysis or dependency modelling. It is used to discover relationship and association rules among variables.
B. **Clustering: -** This technique creates and discovers group of similar data items. This is also called unsupervised classification.
C. **Classification: -** This can classify data according to their classes i.e. put data in single group that belongs to a common class. This is also called supervised classification.

D. **Regression: -** It tries to find a function that model the data with least errors.
E. **Summarization: -** It provides easy to understand and analysis facility through visualization, reports etc . It is possible to mine data with computer that automates this process.

LITERATURE SURVEY

1. **R. B. Oliveira, J. P. Papa, A. S. Pereira, and J. M.R. S. Tavares, "Computational methods for pigmented skin lesion classification: review and future trends,"**

This review provides an overview of current developments of computational methods for skin lesion image classification. Pigmented skin lesion classification is an area of great research interest due to its importance in skin cancer prevention, as well as in the early diagnosis Studies specifically addressing automatic methods applied to the feature selection and extraction steps, based on several clinical approaches, were presented in this review.

In addition, the skin lesion classification step was addressed by including classifiers and evaluation procedures, as well as some performance results for pattern and lesion classification.

2. **J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," Advances in Large Margin Classifiers**

Proposes a system for the automated diagnosis of early melanoma using the ELM 7-point checklist. ELM is the epi luminescence microscopy non-evasive technique that uses different light invasive techniques with an oil immersion technique. The 7 point checklist refers to the a typical network pigment network, blue whitish veil, a typical vascular pattern, irregular streaks, irregular pigmentation, irregular dots and regression structures.

The input of the Computer Aided Diagnostic (CAD) system will be digital images obtained by ELM, which are processed through different algorithms. Then the images are processed in three main stages in which first the boundary detection is done followed by feature extraction where different morphological and chromatic features are considered, followed by classification.

3. **T. DeVries and D. Ramachandram, "Skin lesion classification using deep multi-scale convolutional neural networks,"**

Here the decision tree classifiers belonging to supervised machine learning techniques are used. The decision tree classifier is a predictive model and is preferred as it is fast to train and apply and the rules are easy to understand. Proposes a automatic detection system for melanoma which uses statistical techniques and approaches to improve the performance of different algorithms for automatic detection of dermoscopic criteria provided by 7-point checklist method.

Here, the boundary detection in done by a technique based on adaptive thresholding and also on an unsupervised approach based on statistical region merging. Feature extraction is done by taking into account the first order low level features. These features are measured by techniques like color segmentation which is a statistical region merging technique belonging to the region growing and merging group and texture extraction which is a combination of two different techniques namely structural and spectral methods. Structural technique is intended to search for primitive structures such as lines or points which can constitute a texture. Classification is done by using decision tree classifiers. Spectral technique is based on Fourier analysis of grey level image.

4. **D. Gutmanet al., "Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging (ISBI)**

In his paper proposes a automated method for melanoma diagnosis. The input images are a set of dermoscopic images. The features are extracted based on grey level co-occurrence matrix. Here the classifier used is a multilayer perception classifier which uses 2 different techniques in training and testing process which is the automatic multilayer perception classifier and traditional multilayer perception classifier. This comes under the neural network classifiers.

Results obtained from this method indicate that the texture analysis is a useful method in diagnosis of melanocytic skin tumors with a high level of accuracy.

5. **Razmjooy N, Mousavi BS, Soleymani F, Khotbesara MH (2013) A computer aided diagnosis system for malignant melanomas. Neural Computing and Applications**

Proposes a computer aided diagnosis – a CAD system, which is a decision support system based on semantic analysis of melanoma images. The input isdermoscopic images from Jagiellonian University skin lesions database. The images are then segmented and the objects are extracted which leads to border extraction. The binary border mask is generated and objects are extracted by running simple region growing algorithms. When every object is separated from this image, feature extraction is done as the next step. The colour-based features are extracted and classification is done using classification algorithms likes support vector machines and neural networks.

A support vector machine is used with four kernels namely linear, polynomial, radial and sigmoid. For Neural networks, radial basis function is used. The classification is done for six object groups which are skin regions, red regions, black regions, light and dark brown regions and grey blue regions. Best results of 98.3% are obtained for objects corresponding to dark brown region of images. The second best is the standard skin region which is 97.5%. The classification accuracy for black regions is 93.89% and for pink red regions it is 94.3%.

The least accuracy rate of 80.07% is achieved in the blue grey veil color, because areas covered by them could easily belong to other region types as blue grey veil appear can appear over any other region class. This paper summarizes that the support vector machines with linear kernel prove to perform best in classifying melanoma.

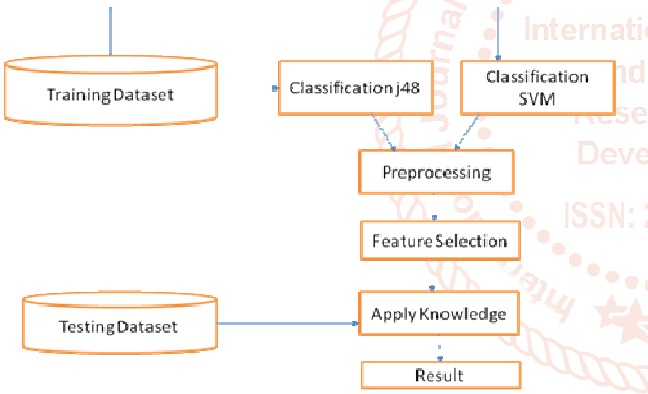
IMPLEMENTATION

In this project, Skin cancer is one of the most common cancer types worldwide. Among different types of skin cancers, skin lesion (the deadliest type) is responsible for 10,000 deaths annually just in the United States. However, if detected early it can be cured through a simple excision while diagnosis at later stages is associated with a greater risk of death-the estimated 5-year survival rate is over 95% for early stage diagnosis, but below 20% for late stage detection. In this existing system of project they are implemented algorithms by separately

.Evaluating a accuracy level of algorithms by individually. So accuracy prediction levels not more accuracy because we have to using single algorithms. we have to worked on more than one algorithm so we can find out which algorithm is better than some algorithm. We have to using real time data set in proposed system.

In this project is using J48 algorithm and SVM algorithm in data mining techniques for lesion skin cancer. System can overcome to diagnosing the disease quickly and accuracy. Comparing to other algorithm j48 algorithm has more accuracy.

It produced more accuracy more than other algorithms. Original dataset can be used in proposed system. Healthy, low risk, moderate risk, high risk is showed to predict the state of disease by analysing the pre-processed data.



Data analysing

Pre-processing: Data pre-processing is an important step in the data mining process. Real world data are generally noisy incomplete, inconsistent. Data cleaning can be applied to remove noisy and correct inconsistencies in data. Data cleaning is typically two step process: First to detect errors in a dataset and then to correct them.

1. Data Cleaning
2. Data Transformation
3. Data Reduction

Features selection methods: Feature selection is the process of finding the meaningful input. It is extracting useful information or features from existing data because Data almost always contains more information than is needed to build the model, or the wrong kind of information. It enables the machine learning algorithm to train faster.It reduces the complexity of a model and makes it easier to interpret. It improves the accuracy of a model if the right subset is chosen. It reduces over fitting.

1. Filter Methods
2. Wrapper Methods

Classification methods:Skin lesion classification using 2 **algorithms**

1. J48,
2. SVM.

The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. The classification step consists of recognizing and interpreting the information about the pigmented skin lesions based on features extracted from images. The classification process generally occurs by randomly dividing the available image samples in training and test sets. The training step consists of developing a classification model to be used by one or more classifiers based on the samples of the training set.

Prediction accuracy: We can use a model to make predictions, or to estimate a dependent variable's value given at least one independent variable's value. Predictions can be valuable even if they are not exactly right. Good predictions are extremely valuable for a wide variety of purposes.

CONCLUSION

Thus the process executed using J48 algorithm and SVM in data mining techniques for selecting the treatment for skin cancer. So that system can overcome to diagnosing the disease quickly and accuracy. Comparing to other algorithm j48 algorithm has more accuracy. Clinical decision support with computer based patient records could reduce medical records. It enhances patient safety. Original dataset can be used in proposed system. Healthy, low risk, moderate risk, high risk is showed to predict the state of disease by analysing the pre-processed data.

FUTURE ENHANCEMENT

Future scope and enhancement in this research we can develop predictive model that can analysis tested data and it will be helpful for medical science and government sector.

References

[1] "Computational Methods for Pigmented Skin Lesion Classification in Images: Review and Future Trends"

[2] "Dermoscopy of pigmented lesions: a valuable tool in the diagnosis of melanoma"

[3] "Dermoscopy Image Analysis: Overview and Future Directions"

[4] "Improving dermoscopy image analysis using color constancy"

[5] "Automated detection and analysis of dermoscopic structures on dermoscopy images"

[6] "Machine Learning Approaches to Multi-Class Human Skin Disease Detection "

[7] " k-nn classifier for skin cancer classification"

[8] "Comparative Analysis of Various Classification Algorithms for Skin Cancer Detection"

[9] "Comparison of Skin Disease Prediction by Feature Selection using Ensemble Data Mining Techniques"

[10] "Analysis on Skin Lesion Classification Systems and Dermoscopic Feature Analysis for Melanoma"