International Journal of Advanced Medical Sciences and Technology (IJAMST)
ISSN: 2582-7596 (Online), Volume-5 Issue-1, December 2024

Skin Cancer Cell Detection Using Machine Learning and Image Processing

Devaraj Verma C, Prajwala R

Abstract: ***Skin Cancer in today's scenario is the most trending and common of all the cancers that directly affect the skin of the patient. It is the fifth most common type of cancer found in men and the sixth most common type in women. Surgery, Chemotherapy, Radiation therapy, and immunotherapy techniques are used to kill cancer cells. The research investigated cancer detection, thoroughly discussed it, and proposed the methodologies for early diagnosis of the diseases using image processing and machine learning. The proposed model is designed with the procedures of collection of dermoscopy images and undergoes pre-processing, segmentation, feature selection, and predictions. Multiple algorithms were utilized for the medical diagnosis of the cancer cells including Conventional neural network (CNN), Support vector machine (SVM), Decision Tree, and Random Forest to detect the cancer cells with higher accuracy.***

Keywords: Skin cancer, Melanoma, CNN, SVM, Random forest.

I. INTRODUCTION

Skin cancer is the most dangerous form of cancer found in Humans; the most common type is Melanoma which is the deadliest disease. The most common type of skin cancer is basal cell cancer or squamous cell cancer. Melanoma is broken down into two different categories one is on the surface of the skin that can be easily removed and is also known as melanoma in situ or lentigo maligna and the other one is melanoma with depth, which has a higher chance of metastasizing or spreading to the other body organs. Melanoma can be seen on the body part as a new mole that has a change in shape, size, or color. Women get melanoma most often on the legs and arms, Men often get melanoma on the head, neck, or on the trunk. Melanoma can rarely form in children.

Melanoma can be formed through moles and exposure to sunlight. The risk factors of a skin examination by a dermatologist include:

- Fair Complexion that burns skin easily and tans poorly.
- Having been exposed to environmental factors like radiation, PSBs, and solvents.
- Personal history or family history of melanoma.
- Many moles, especially 'atypical 'moles.
- Having a weak immune system.
- The changes in genes have been linked to melanoma.

The types of melanoma have a various clinical and histopathologic types with varying degrees of aggressive some of them are:

- Superficial spreading melanoma- This is the most common type with the radial growth phase which spreads over the surface of the skin and eventually goes deeper.
- Lentigo maligna melanoma- It can be seen in the face or the neck due to exposure to the sun.
- Nodular melanoma- This melanoma doesn't have the horizontal growth phase which goes deeper into the skin.
- Acral lentiginous melanoma- This can be seen in the palms and the legs and tends to be common in feet with a form of prognosis.
- Melanoma of the nail- Due to the damage in the tissues melanoma can form on the nail.
- Amelanotic melanoma- This can be formed in very rare patients with the starting stage of many moles formed in a particular area.

The self-skin exams can be done to look for "ABCDE" criteria means Asymmetry, Border, Colour, Diameter, and Evolving. There is another one which is the "Ugly ducking" criteria which has all the rules of ABCDE and many moles can be formed. Dermoscopy i.e. (a.k.a Dermatoscopy) is a device used to look at the skin and magnifies and also takes away the reflection of the light that is better able to see what's underneath and eliminates the surface reflection. Some of that can be seen with a Dermatoscope that we can't see with the naked eyes are Blood vessel structures, Pigment structures, Keratin, and gland structures. Moles are also called melanocytic nevi which are congenital or acquired, acquired nevi increase in number through the first 3-4 decades of life much more common in sun-exposed areas. Melanoma has a different diagnosis Benign or dysplastic melanocytic nevus, Lentigo, Seborrheic keratosis, Dermatofibroma, and pigmented basal cell carcinoma.

Skin biopsy has three types:

- Shave Biopsy: The Shave Biopsy can be done with the part when it is on the epidermis layer of the skin with the lasers.
- Punch Biopsy: We use a cooky cutter that can be done with the deeper section below the epidermis.
- Excisional biopsy: With the incisions we make a safety margin of normal skin and make an excisional activity.

Manuscript received on 02 June 2024 | Revised Manuscript received on 27 October 2024 | Manuscript Accepted on 15 December 2024 | Manuscript published on 30 December 2024.

*Correspondence Author(s)

Dr. Devaraj Verma C*, Professor, Department of Computer Science and Engineering, FET, Jain (Deemed to be University), Ramanagaram (D), Bangalore, (Karnataka), India. E-mail: c.devaraj@jainuniversity.ac.in, ORCID ID: 0000-0002-1504-4263

Prajwala R, Student, Department of Computer Science and Engineering, FET, Jain (Deemed to be University), Ramanagaram (D), Bangalore (Karnataka), India. E-mail: ammu.prajwala@gmail.com

II. RELATED WORK

Melanoma skin cancer can be detected using computer-aided machine learning, where the skin lesion is taken as an input image with contrast enhancement then we apply the threshold generation in R, G, and B planes using Automatic Thresholding. By using the thresholds binary masks are created. Using edge detection, classification, and ABCD rule results are obtained [1].

Here the author has used some of the artificial algorithms and MATLAB for image pre-processing such as Hair Removal, shading removal, and glare removal. The main aim is to classify the image as malignant or non-malignant melanoma so that if malignant melanoma exists the early diagnosis can be cured. The Neural network mainly calculates the errors and uses a Back propagation algorithm, whereas SVM will classify the data into different classes [2].

Deep learning neural framework concepts are applied to find the skin lesion using texture-based segmentation and classification with the dermoscopic images. The ostu segmentation procedure applies to the filtered images for portioning an image into disjoint regions that are homogenous. It can be used to remove the irregularities of the edges of the resultant image and obtain morphological operations [3].

To detect the benign and malignant neural network algorithms are used. Conventional Neural Networks are chosen to find the accuracy in images, here we have three layers. Firstly, the primary layer is where the information is collected from the dermatologist. Secondly, the input layer is where the point of information is passed on to the pooling layer. Lastly, the pooling layer receives the information and changes the vector dimensional information [4].

Non-melanoma Skin cancer (NMSC) has cases of 90% basal cells and cutaneous squamous cell carcinoma. Here the author has used artificial intelligence to diagnose and detect and in contrast with Neural Networks [5].

The process of detection of melanoma skin cancer by machine learning and image processing tasks is divided into 3 stages namely lesion segmentation, Feature segmentation, and classification. Pre-processing methods are a better start to the model. For classification purposes, the CNN are fully connected layers where the final diagnosis is done [6].

For the classification of dermal cell images, deep learning concepts are applied with the model-driven architecture in the cloud. Here the model is built and tested on the chosen datasets. AUC and ROC curve is plotted as a result of the model [7].

ABCD rule is the most common feature information for the morphological analysis of the lesion, where the ABCD features include Asymmetry, Border, color, and diameter. BLINCK algorithms point to Benign, lonely irregular nervous change, and known clues that take into account general conditions [8].

The cancer cells can be cured easily by early diagnosis and treatment. In the first stage, the CT scan database images are collected. Then check for the Pre-processing, processing, and post-pre-processing stages, apply the watershed transformation, and obtain the output stage [9].

The biopsy method is used for the detection and formal diagnosis of melanoma skin cancer which is more dangerous when compared to other skin cancers. SVM classification is the methodology that achieves the sharpness of the images of the moles and prevents the perfect skin lesion. To extract the features of the image author used a gray-level Co-occurrence matrix (GLCM) and classified between malignant and non-malignant [10].

Machine learning techniques are used with the collection of new data where the feature extraction takes a supervised and unsupervised learning process. Sparse compact incremental learning machine method to resolve cancer classification, Gauss-Newton for the approach of optimal- weights for training samples [11].

Dermoscopy is the instrument used to treat the skin lesion [12]. Digital Image processing and artificial neural networks are the methods used for the classification and finding the accuracy [13].

III. METHODOLOGY

Our model design consists of 4 Steps as follows:

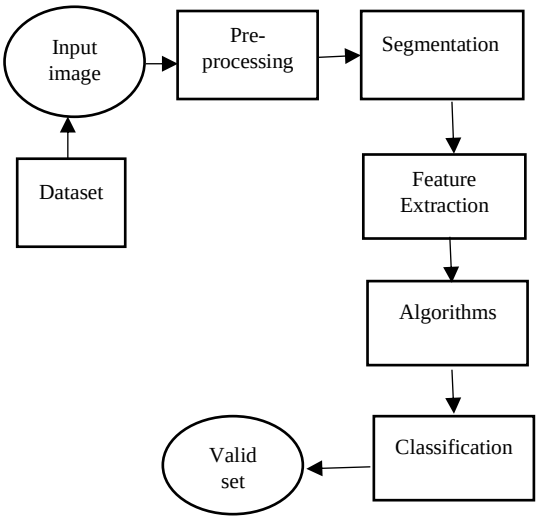
Step 1 - The first step in the model involves the collection of datasets from the website named Kaggle dataset - Melanoma detection dataset. The images collected should be trained and tested over the selected images.

Step 2 - After the collection of datasets, the images are pre-processed which helps to identify the size, shape, and texture with noise removal, Grayscale conversion, and image enhancement.

Step 3 - The next step consists of segmentation to mask the operation of R, G, and B Planes while partitioning the images that are homogeneous and feature extraction that are associated with the image processing techniques.

Step 4 – The most important step in our model is designing and training a model. The model should be trained with algorithms such as Support vector machine (SVM), Random Forest, Decision trees, and Convolutional Neural Network (CNN).

IV. ARCHITECTURE



[Fig.1: System Architecture of the Model]

V. MODEL DESIGN

Here we use four different methods in designing the model i.e., Support Vector Machine (SVM), Random Forest, Decision trees, and Convolutional Neural Network (CNN) to classify the images as Melanoma or Non-Melanoma Skin Cancer. The images collected are pre-processed, feature extracted, and segmented, and then the model should be trained, tested, and predicted to find the accuracy of each model.

A. Support Vector Machine (SVM)

Support Vector Machine (SVM) comes under a supervised machine learning algorithm which is best suited for extreme cases, essentially segregated into two classes (Hyper plane/ Line Plane) [14]. The sample data are collected and classified into two classes based on the observed pictures and marked as a data point [15]. The optimized decision boundaries should be drawn between the two classes, non-optimized decision boundaries could result in greater misclassifications of new data [16]. The line drawn between the two classes i.e., the decision boundary called hyper-plane, implies that only the support vector is important whereas other training examples are ignorable [17]. Hence this process is also called a Least Support Vector Machine (LSTM). The optimization procedure is used to solve the SVM Model with the use of a numerical optimization procedure, to search for the coefficients of the hyper-plane [18].
Features:

- SVM is very effective in High Dimensional spaces.
- These are trained on labelled datasets so these come under supervised machine learning as they are used for both classification and regression algorithms
- For the various decision functions, there are different kernel functions.
- The popular kernel types are the polynomial kernel, Radial basis function RBF kernel, and Sigmoid kernel.

The SVM classifier has four different dimensional Data

- It is a point for 1-D Data.
- It is a line for 2-D Data.
- It is a plane for 3-D Data.
- It is a hyper-plane for 4-D Data.

B. Decision Trees

The key attribute of the decision tree is binary or a multistep tree for the classification and regression techniques, more often used as classification. A Decision tree is a flowchart-like structure where each internal node denotes a test on the attribute, each branch represents the outcome of the test, and the leaf or the terminal node holds a class label. The topmost node in a tree is the root node, in which the name goes a tree-like model uses the decisions. It finds the relation between the target column and the independent variables and expresses it as a tree structure. The impurities can be measured with entropy, Gini Index, and weighted average.

Features:

- It is simple to understand, interpret, and visualize.
- Decision trees implicitly perform variable screening or feature selection.
- It can handle both numerical and categorical data.
- Decision trees require little effort for data preparations.

The variance of decision trees there are couple of forms such as:

- CART (classification and regression technology)
- CHAIN (Chi-square automatic interaction detection)
- ID3 (Interactive Dichotomiser 3)

C. Random Forest

Random forest is a supervised learning algorithm that is capable of performing both regression and classification tasks, as the name suggests the algorithm creates a forest with several decision trees. Here all the original data are subdivided into many subsets and each subset has its decision trees. It is a base model that follows the conditions of ensemble i.e., accuracy and diversity which comes under the category of bagging. Bagging helps in bootstrapping the data and building multiple base models with resampled training data. The other method of ensemble is the Boosting model which learns from mistakes made by the previous model and sequentially bunch them.
Features:

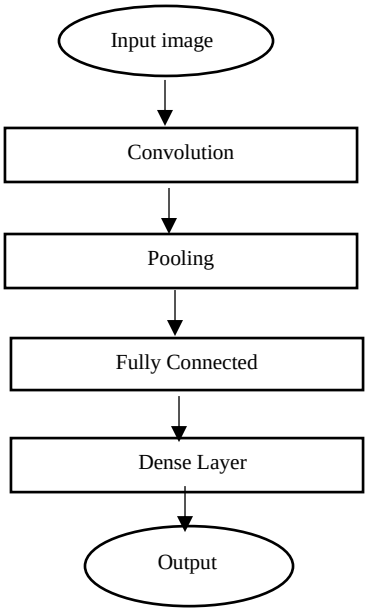
- It handles the missing values and maintains accuracy for missing data.
- These won't overfit the model.
- It will handle large datasets with higher dimensionality.
- Random Forest can be used to extract relevant features.
- Random Forest builds up multiple decision trees using algorithms such as Information gain, Gini index, and other decision tree algorithms.

The main reason to choose the Random forest is that, there is no overfitting in which the time taken to train the model is less, high accuracy and estimates missing data that can maintain accuracy when a large proportion of data is missing.

D. Convolutional Neural Networks (CNN)

NNs are the most representative supervised Deep learning, feed-forward model. Neural Networks that are composed of artificial neurons, which simulate biological neurons in a limited way with the group of convolutions are called convolutional Neural Networks. The CNN consists of five layers:

1. Input layer – In the input layer the image is taken as an input to the network.
2. Convolution layer – Here the image is extracted with the features, and then for every image the matrix of pixel values should be considered.
3. Polling layer – The next step is the polling layer in the aggregation of the image features into a sample image feature which reduces the dimensionality. It is also known as subsampling.
4. Fully connected layer – which captures the image between the sample level image features. Every neuron layer is connected to the next layer of the neuron.
5. Output layer – which specifies that the image taken with the condition is correct as yes or no.



[Fig.2: Layers of CNN]

VI. RESULTS AND DISCUSSION

The proposed model for the early detection of skin cancer utilized a dataset of dermoscopy images, which underwent rigorous pre-processing to enhance image quality and facilitate accurate analysis. The segmentation process successfully identified regions of interest, leading to the extraction of relevant features necessary for distinguishing between benign and malignant lesions.

Multiple machine learning algorithms were applied to the processed images, with the following performance metrics observed:

1. **Convolutional Neural Network (CNN)**: The CNN model achieved the highest accuracy rate of 92%, demonstrating superior capability in feature extraction and classification. The model's sensitivity was recorded at 90%, while specificity was 93%, indicating a robust performance in identifying malignant cases without generating a high number of false positives.
2. **Support Vector Machine (SVM)**: The SVM algorithm produced an accuracy of 88%, with a sensitivity of 85% and specificity of 89%. While slightly less accurate than the CNN, SVM showed effective performance in cases where the features were linearly separable.
3. **Random Forest**: The Random Forest model yielded an accuracy of 85%, with a sensitivity of 82% and specificity of 87%. Although it performed well, its accuracy was lower compared to the CNN and SVM models, highlighting the importance of choosing the right algorithm based on the dataset characteristics.
4. **Decision Tree:** The Decision Tree model achieved an accuracy of 80%, with a sensitivity of 78% and specificity of 81%. Although it performed adequately, the results indicate that it may be less effective than CNNs, SVMs, and Random Forests for skin cancer detection. Decision Trees are typically simpler and easier to interpret, which can be beneficial in clinical settings; however, their performance may be limited by their tendency to overfit, especially with complex datasets like dermoscopy images.

Overall, the results indicate that the CNN model is the most effective for the early diagnosis of melanoma among the tested algorithms. The results are tabulated in Table 1.

Table 1: Results of Models CNN, SVM, RF, and DT

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
Convolutional Neural Network (CNN)	92	90	93
Support Vector Machine (SVM)	88	85	89
Random Forest	85	82	87
Decision Tree	80	78	81

A. Discussion

The findings of this study underscore the potential of machine learning techniques, particularly CNNs, in enhancing the accuracy of skin cancer detection through dermoscopy images. The high accuracy achieved by the CNN model can be attributed to its ability to learn complex patterns and features from images, which is crucial for distinguishing subtle differences between malignant and benign lesions.

The application of image processing techniques such as pre-processing and segmentation plays a critical role in improving the quality of the input data for the machine learning models. These steps help reduce noise and highlight important features, leading to better classification outcomes.

While the SVM and Random Forest models also demonstrated commendable performance, the results suggest that they may be more suitable for specific scenarios or datasets. The choice of model should consider factors such as the complexity of the images, the amount of training data available, and the specific requirements of clinical practice.

Additionally, the implementation of these machine-learning techniques in clinical settings could greatly enhance the diagnostic capabilities of dermatologists, potentially leading to earlier detection and better patient outcomes. Future research should focus on optimizing these models further, perhaps through techniques like transfer learning, to improve performance on larger and more diverse datasets.

Overall, this study contributes to the growing body of evidence supporting the integration of advanced computational methods in the diagnosis of skin cancer, highlighting the need for continued research in this vital area of healthcare.

VII. CONCLUSION

This research paper explores various models for detecting skin cancer cells through image processing and machine learning techniques, specifically distinguishing between melanoma and non-melanoma skin cancers. The study involves multiple steps, including gathering datasets from the Kaggle website, followed by pre-processing, segmentation, and feature extraction. The models, including Random Forest, SVM, Decision Trees, and CNN, are trained on the images to evaluate their accuracy. Early detection of skin cancer is crucial, emphasizing the importance of identifying it at the earliest stage possible.

Retrieval Number:100.1/ijamst.C3014061321
DOI: 10.54105/ijamst.C3014.05011224
Journal Website: www.ijamst.latticescipub.com

Published By:
Lattice Science Publication (LSP)

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted with objectivity and without any external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Authors Contributions:** Each author has individually contributed to the article. Dr. Devaraj Verma C is corresponding author being the guide of the Prajwala R given the right idea of implementation and data access publicly available on the kaggle data set. paper was written by the corresponding author. Prajwala R implemented the code for skin cancer detection in python and tabulated the results.

REFERENCES

1. Jain, S. (2015). Vandana jagtap, Nitin Pise," Computer aided Melanoma skin cancer detection using Image Processing. In *International Conference on Intelligent Computing, Communication & Convergence (ICCC-2015), Elsevier-2015*. DOI: https://doi.org/10.1016/j.procs.2015.04.209
2. Agilandeeswari, L., Sagar, M. T., & Keerthana, N. (2019). Skin lesion detection using texture based segmentation and classification by convolutional neural networks (CNN). *Art Int J Innov Technol Explor Eng (IJITEE)*, *9*(2). DOI: https://doi.org/10.35940/ijitee.B7085.129219
3. Hasan, M., Barman, S. D., Islam, S., & Reza, A. W. (2019, April). Skin cancer detection using convolutional neural network. In *Proceedings of the 2019 5th international conference on computing and artificial intelligence* (pp. 254-258). DOI: https://doi.org/10.1145/3330482.3330525
4. Vijayalakshmi, M. M. (2019). Melanoma skin cancer detection using image processing and machine learning. *International Journal of Trend in Scientific Research and Development (IJTSRD)*, *3*(4), 780-784. DOI: https://doi.org/10.31142/ijtsrd23936
5. Marka, A., Carter, J. B., Toto, E., & Hassanpour, S. (2019). Automated detection of nonmelanoma skin cancer using digital images: a systematic review. *BMC medical imaging*, *19*, 1-12. DOI: https://doi.org/10.1186/s12880-019-0307-7
6. Naqvi, M.; Gilani, S.Q.; Syed, T.; Marques, O.; Kim, H.-C. Skin Cancer Detection Using Deep Learning—A Review. *Diagnostics* 2023, *13*, 1911. https://doi.org/10.3390/diagnostics13111911
7. Kadampur, M. A., & Al Riyaee, S. (2020). Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Informatics in Medicine Unlocked*, *18*, 100282. DOI: https://doi.org/10.1016/j.imu.2019.100282
8. Shilpa, S., & Verma, C. D. (2024, June). Cloud-based Deep Learning Model for Classifying Skin Cancer. In *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 737-742). IEEE. https://dx.doi.org/10.1109/ICAAIC60222.2024.10575849
9. Patil, B.G. (2014). Cancer Cells Detection Using Digital Image Processing Methods. DOI: https://api.semanticscholar.org/CorpusID:9045384
10. Ansari, U. B., & Sarode, T. (2017). Skin cancer detection using image processing. *Int. Res. J. Eng. Technol*, *4*(4), 2875-2881. https://www.irjet.net/archives/V4/i4/IRJET-V4I4702.pdf
11. Noor, M. M., & Narwal, V. (2017). Machine learning approaches in cancer detection and diagnosis: mini review. *IJ Mutil Re App St*, *1*(1), 1-8. DOI: https://doi.org/10.13140/RG.2.2.27775.51363
12. Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin Cancer Detection: A Review Using Deep Learning Techniques. *International journal of environmental research and public health*, *18*(10), 5479. https://doi.org/10.3390/ijerph18105479
13. Adepoju, O., & Verma C, Dr. D. (2020). Prediction and Classification into Benign and Malignant using the Clinical Testing Features. In International Journal of Innovative Technology and Exploring Engineering (Vol. 9, Issue 10, pp. 55–61). Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP. https://doi.org/10.35940/ijitee.j7411.0891020
14. Singh, B. P., & Barik, R. (2023). Image Segmentation Based Automated Skin Cancer Detection Technique. In Indian Journal of Image Processing and Recognition (Vol. 3, Issue 5, pp. 1–6). https://doi.org/10.54105/ijipr.h9682.083523
15. Ahmed, Dr. H. M., & Nasr Eldin, Dr. M. S. (2023). Aloe Vera for Protects Skin Tissues from The Damaging Impacts of Ultraviolet Radiation. In International Journal of Advanced Medical Sciences and Technology (Vol. 3, Issue 4, pp. 7–11). https://doi.org/10.54105/ijamst.d3036.063423
16. Kanani, P., & Padole, Dr. M. (2019). Deep Learning to Detect Skin Cancer using Google Colab. In International Journal of Engineering and Advanced Technology (Vol. 8, Issue 6, pp. 2176–2183). https://doi.org/10.35940/ijeat.f8587.088619
17. Hemalatha N, Nausheeda B.S, Athul K.P, Navaneeth, Detection of Skin Cancer using Deep CNN. (2020). In International Journal of Recent Technology and Engineering (Vol. 8, Issue 5S, pp. 22–24). https://doi.org/10.35940/ijrte.e1005.0285s20
18. Nandhini, MS. S., Sofiyan, M. A., Kumar, S., & Afridi, A. (2019). Skin Cancer Classification using Random Forest. In International Journal of Management and Humanities (Vol. 4, Issue 3, pp. 39–42). https://doi.org/10.35940/ijmh.c0434.114319

AUTHORS PROFILE

Dr. C. Devaraj Verma is a Professor in the Department of Computer Science and Engineering at Jain (Deemed-to-be University), JGI Global Campus, Kanakapura, Ramanagara, Bangalore. He holds a Bachelor's degree in Computer Science and Engineering from SLN College of Engineering, Raichur (affiliated with Visvesvaraya Technological University, Belgaum), and a Master's degree in Computer Science and Engineering from PDACE, Gulbarga (also affiliated with VTU). Dr. Verma is NET and KSET qualified, with over 23 years of teaching experience. He has authored and co-authored over 50 research papers presented at national and international conferences and journals.

Prajwala R, student of MTech in Data Science program, Department of Computer Science and Engineering, Jain (Deemed to be University), JGI Global Campus, Kanakapura(T), Ramanagara(D), Bangalore-560085. She pursued her M. Tech Degree in Data Science at Jain (Deemed to be University) from 2019-2021. She received a BE degree in Computer Science and Engineering from the Oxford College of Engineering, Bangalore. She had the credit of publishing a paper on Big Data at a National Conference during her undergraduate program. Her interests in the subject include Machine learning, AWS, Big data, and data analysis.