



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Diseases Classification using advance Machine learning Techniques

Manju Payal¹, K. Suresh kumar², Dr. T. Ananth kumar³

Software Developer, Academic Hub Ajmer, India¹

Department of Electronics and Communication Engineering, IFET College of Engineering, Tamilnadu, India²

Department of Computer Science and Engineering, IFET College of Engineering, Tamilnadu, India³

ABSTRACT: Machine learning algorithms are being utilised in the processes of screening patients, making diagnoses, determining prognoses, and estimating therapeutic responses. All of these tasks are currently being performed. The principles of machine learning are implemented in a significant way across the bioinformatics research community. Machine learning is a subset of artificial intelligence that is used to develop strategies based on historical data correlations and data patterns. This is accomplished through the use of data pattern recognition and data pattern recognition software. This is achieved with the assistance of data pattern recognition and software designed specifically for data pattern recognition. Among the many areas that benefit from the implementation of machine learning are bioinformatics, the search for new medications, personalised treatment, and skin cancer. These applications are not exclusive to these areas, however.

KEYWORDS: Machine learning; Healthcare; Bio Medicine; Prediction.

I. INTRODUCTION

The field of healthcare all over the world is undergoing rapid expansion. According to a recent study, in spite of the fact that the economy is growing more slowly and there are concerns regarding global trade, it is likely that this will exceed \$2 trillion this year. The average lifespan of humans is increasing, as is their general level of health [1]. Living organ donation is gaining popularity. Robots are utilised in the surgical removal of gallbladders, replacement of hips, and transplantation of kidneys. It is now possible to detect skin cancers early on and with a minimal amount of room for error on the part of medical professionals [2]. Patients who have undergone reconstructive surgery for breast cancer have been able to re-grow their glands [3]. In spite of this, it is anticipated that the current global population of 7.7 billion will increase to 8.5 billion by the year 2030, which may make it difficult to satisfy the demand for health care [4-8]. Significant advances in machine learning (ML) can help anticipate infection risks, improve diagnostic accuracy, and provide individualised treatment options in this field [9]. The COVID-19 epidemic is causing significant economic disruption across a variety of industries, which is having a knock-on effect on the day-to-day activities of a large number of local residents[10]. In order to determine the best strategies for optimal commercialization, small and medium businesses in certain industries need to be informed not only of the progression of the pandemic but also of the feelings associated with the affected customers. This paper proposes the use of an expert system that is built on a combination of machine learning and sentiment analysis in order to help businesses make decisions involving the fusion of data obtained through web scraping. Using an artificial intelligence system that is centred on humans, the technology can automatically generate explanations. The expert system receives web content from a variety of sources through the utilisation of a scraping module.

Machine learning is being implemented in the healthcare industry. In 2004, it was established in the province of Ontario with the purpose of determining which patients had a high probability of being readmitted or passing away within the first month after being discharged from the hospital. The algorithm takes into account four different aspects: the duration of the patient's hospital stay, the severity of their condition upon admission, any concurrent illnesses, and visits to the emergency room[6]. The LACE measure, which is renowned for being founded on the theory of machine learning, is widely recognised as a barometer of the quality of care. The previous medical histories of the patients are consulted in order to make accurate projections regarding the patients' future health[7]. It makes it possible for medical professionals to disperse resources in a timely manner in an effort to reduce the number of deaths. This technology is currently being used by the Knight Cancer Institute in Oregon and Microsoft's Project Hanover to personalise drug combinations in an effort to cure blood cancer. This technology is being used

to do everything from the initial screening of drug compounds to the calculation of the success rates of a specific medicine based on the physiological factors of the patients[9].

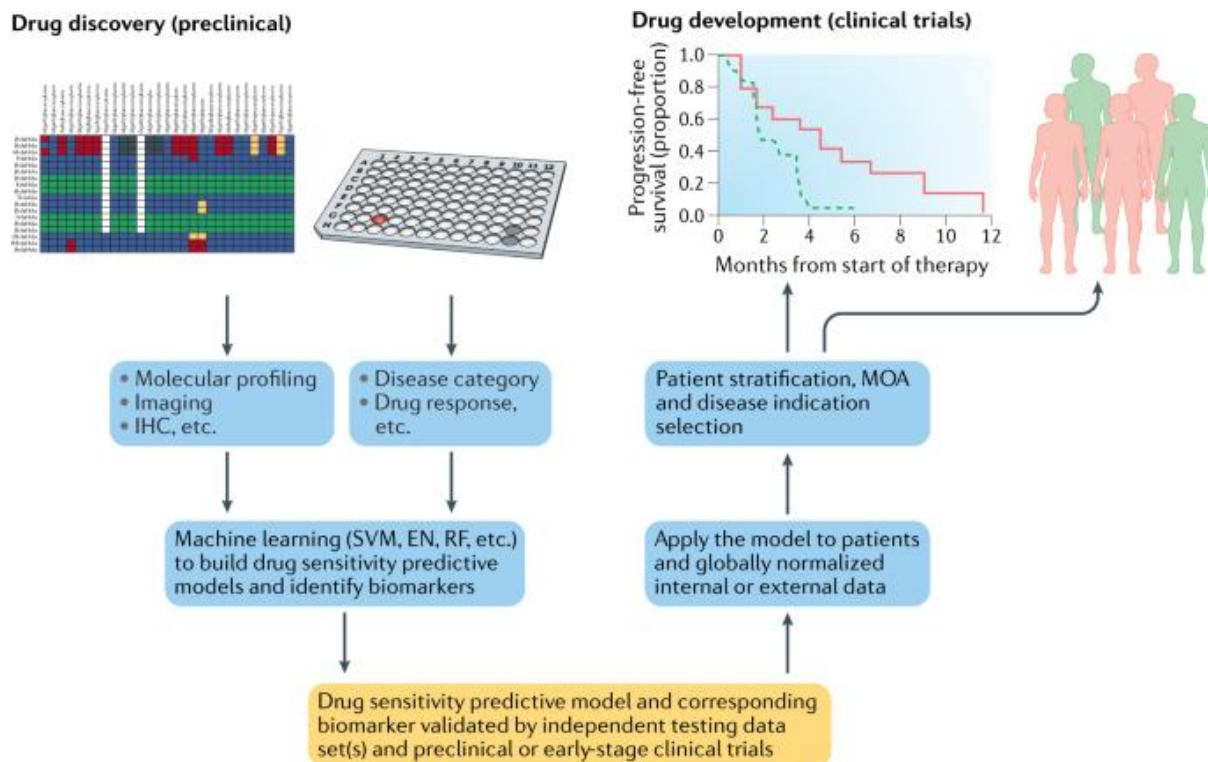


Fig.1. Machine learning based Drug discovery and drug development model

ML approaches provide a set of tools that can improve discovery and decision-making in situations where there is a significant amount of high-quality data and well-defined issues. The application of machine learning is useful throughout the entirety of the drug development process. Target validation, the identification of prognostic biomarkers, and the evaluation of digital pathology data in clinical trials are just a few examples of the many applications of digital pathology.

II. RELATED WORKS

Precision medicine and next-generation sequencing are two emerging technologies that both make use of machine learning to improve the likelihood that patients will experience the therapeutic benefits of a given medication. In the field of medication research, numerous machine learning strategies have been implemented, including Random Forest (RF), Naive Bayesian (NB), and support vector machine (SVM). [12]

a. **Support Vector Machine (SVM):** SVMs are supervised machine learning algorithms that are used in the drug development process to identify classes of drugs. SVMs do this by deriving a hyperplane from the feature selection.

[13] It does this by drawing on the similarities between classes in order to produce an infinite number of hyperplanes. It prepares itself for linear data by categorising chemicals according to a set of predetermined characteristics and then mapping those classes into a chemical feature space. It is represented by a hyperplane, which functions as a tool for classifying data points through the establishment of decision boundaries. Maximizing the margin between categories in N-dimensional space results in the formation of an ideal hyperplane (N is the number of features). [14] In the process of drug discovery, support vector machines (SVM) play an essential role due to their ability to differentiate between active and inactive compounds, rank compounds obtained from each database, and train regression models. [41] Regression models are essential to comprehending the connection between the medication and the ligand as a result of the fact that the forecasting process involves a query for the datasets. When screening a large number of active compounds against a single protein of interest, SVM can be applied in a number of different contexts. [15] In the

context of SVM classification, a subset binary class prediction is able to differentiate between active and inactive molecules.

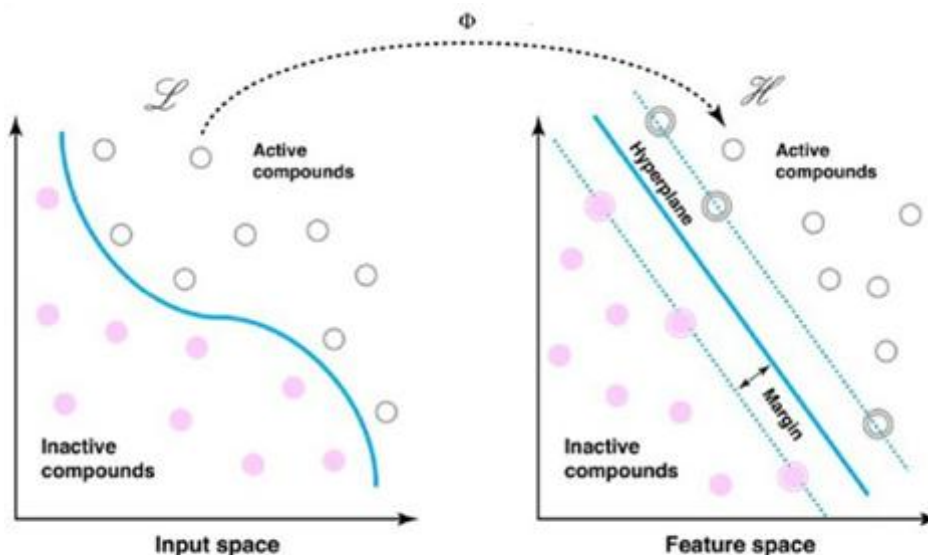


Fig. 2. Active SVM model of space in machine learning

It is possible for it to rank compounds from different databases according to the likelihood that they will be active in any computational screening for drug discovery. SVM can be utilised in a number of different ways to obtain results, with the primary objective being to differentiate between substances that are active and those that are inactive[16]. In order to train the algorithm, the process could be modified by making use of various descriptors for feature selectors. Some examples of these would be 2D fingerprints and target protein. Compounds are able to be ranked from the most selective to the least selective based on the direction the chemical is located in relation to the hyperplane, which results in the creation of a negative or positive class label for the chemical. On the other hand, kernel functions are utilised in order to optimise the results when dealing with non-linear data. Data are plotted using kernel functions in a higher-dimensional space that allows for the possibility of class separation.

b. Random Forest (RF) Classifier

In the process of drug discovery, RFs are utilised for the most part for feature selection, classifiers, and regression. Through the inclusion of information regarding genomic sequencing, RF was able to contribute to the maintenance of error and obtain medication responses based on genomic characterizations. Multivariate RFs are intended to cut down on error by employing a wide range of different methods for error estimation within the system itself [19]. The genetic and epigenetic characterisation combinations of the data are fed into the computational framework, which then predicts the mean and confidence interval of medication reactions based on the data. A fundamental requirement for the evaluation of any drug that will be put through clinical testing [18]. In an effort to make more accurate predictions based on the response profile, The authors combined the modelling framework with functional RF in their research. They made an effort to address the difficulties that individuals have experienced in acquiring the appropriate chemicals for their cancers.

RF was utilised in the process of generating the node and leaf nodes for the regression tree[21]. It was able to gather data points relating to dose-response relationships. The predictions regarding the dose-response profile are generated by the leaf nodes of the algorithms, which are also responsible for storing the functional data. The genomic sequences and the characteristics of those sequences are what are recorded as the model's data [20]. RF algorithms have also been used as a classification and regression tool in quantitative structure-activity relationship (QSAR) modelling, which is part of the lead discovery process.

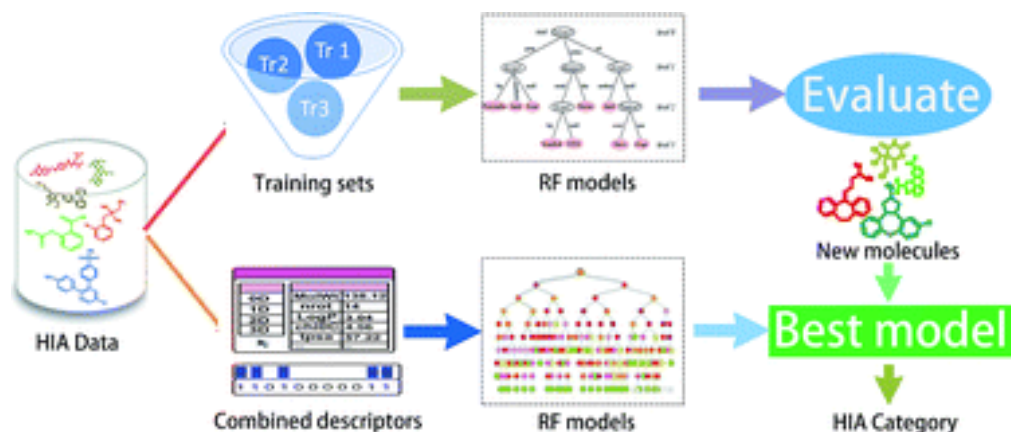


Fig. 2. Random Forest Classifier based training and prediction model

RF is a common algorithm that is designed to remove outliers from large datasets along with classifying and designing them based on relative characteristics that are defined for the method. It was developed specifically for use with datasets that contain a wide variety of features [22]. Medical professionals can now, for example, use algorithms to better understand the processes of disease and develop novel treatments for conditions like type 2 diabetes[23].

1. Optimized Patient Recruitment for Clinical Trials

Finding willing participants for clinical studies can be challenging. A number of criteria need to be implemented in order to select participants who are appropriate for the study. Machine learning makes it easy to collect patient data, including historical medical records, psychological behaviour, and the medical history of the patient's family, among other types of information. [24]

In addition to this, the technology is utilised to monitor the biological measurements of the participants in the clinical trials as well as the potential effects that the trials may have in the long run. With such convincing evidence, medical professionals are able to reduce the length of the trial, which in turn lowers the overall costs and increases the efficiency of the experiment. [25]

1.1. Personalized Treatments

Every human body functions in its own special way. When it comes to the same meal, treatment, or season, various people will have wildly divergent responses. This leads to allergic reactions in people. If this is the case, then why is it that the idea of personalising treatment options for a patient based on that patient's medical data is still such a foreign one? [26]

Machine learning provides assistance to medical professionals in determining the risk for each individual patient based on the patient's symptoms, historical medical records, and family history. These factors are collected using micro-bio sensors. These minuscule devices keep track of a patient's health and identify irregularities without favouritism, which enables more advanced capabilities for monitoring their health. [27]

1.2. Early Skin Cancer Detection

According to Cisco, the number of machine-to-machine connections in the global healthcare industry is increasing at a compound annual growth rate (CAGR) of 30 percent, making it the industry with the highest rate of any sector.

The primary application of machine learning is the collection and analysis of patient data in order to identify patterns, which can then be used to diagnose a variety of medical conditions, including skin cancer. More than 5.4 million people in the United States are diagnosed with this condition every single year. The process of diagnosis is one that must be performed virtually and can be very time consuming. Patients go through extensive clinical screenings, which typically consist of a biopsy, dermoscopy, and histopathological examination.

When detected at an earlier stage, skin cancer has a significantly better prognosis. After that, it makes the distinction between benign and malignant lesions by using the experience it has gained along the way and its own set of criteria. Following this step, the analysis will determine a risk assessment score for both melanocytic and non-melanocytic skin lesions, which will enable the doctors to confirm their diagnoses.

According to the authors, "Moleanalyzer pro includes the ability to manually evaluate lesions using recognised checklists in addition to an AI-based unique methodology that allows for a risk-of-malignancy assessment." "Moleanalyzer pro includes the ability to manually evaluate lesions using recognised checklists" During the course of the past few years, a substantial number of dermoscopy images have been used to educate the newly developed algorithm. In order to train its algorithm, FotoFinder Systems relies on a global network of partners who are willing to contribute images of pathologies that have been histologically confirmed. [36]

According to the authors, a research study that included 58 dermatologists from 17 different countries indicated that while professionals accurately recognised 86.6 percent of dangerous skin cancers, Moleanalyzer pro correctly discovered 95 percent of them. While experts were only able to correctly identify 71.3 percent of benign naevi, the technology was able to correctly recognise 82.5 percent of them. [35]

According to the authors, "The MI represents a 'silent virtual colleague' who offers a virtual opinion in a straightforward, uncomplicated manner and at any given time." "However, when combined with the human experience offered by the complementary second opinion service, the diagnostic accuracy of the tool is significantly improved." [34]

Oncologists are able to confirm a diagnosis of skin cancer using evaluation techniques in conjunction with ML, which enables them to begin treatment earlier than is typically the case. Moleanalyzer was successful in identifying malignant skin lesions 96% of the time, whereas professionals were only successful in doing so 86.66% of the time. [31]

1.3. Clinical Performance

It is in the best interest of the government for healthcare providers to submit reports that contain copies of all pertinent patient records that were managed at their facilities.

It is more important than ever before to ensure that hospital locations are compliant with all applicable laws and are operating within the appropriate parameters. This is due to the fact that compliance laws are always being updated. The process of collecting data from a wide variety of sources, correctly arranging it, and employing a number of different methodologies is greatly simplified by machine learning. "It could be a time-consuming task for data administrators to compare patient data from a variety of clinics in order to ensure compliance with the regulations. The researchers explain how machine learning can be used to obtain, compare, and maintain data in a manner that is compliant with government regulations.

III. CONCLUSION

It frees up time for medical professionals so that they can concentrate on providing care for patients rather than looking for or entering data. The second primary goal of machine learning applications in the healthcare industry is to achieve greater diagnostic precision. On the other hand, machine learning completely transforms everything. Moleanalyzer is a piece of artificial intelligence software that was developed in Australia. Its primary functions are to estimate and compare the size, diameter, and structure of molecules. It gives the user the ability to take pictures at predetermined intervals so that they can better distinguish between benign and cancerous skin lesions.

REFERENCES

1. Manju Bala, P., S. Usharani, T. Ananth Kumar, R. Rajmohan, and M. Pavithra. "Blockchain-Based IoT Architecture for Software-Defined Networking." In *Blockchain, Artificial Intelligence, and the Internet of Things*, pp. 91-115. Springer, Cham, 2022.
2. El Hussein, S., Chen, P., Medeiros, L. J., Wistuba, I. I., Jaffray, D., Wu, J., & Khoury, J. D. (2022). Artificial intelligence strategy integrating morphologic and architectural biomarkers provides robust diagnostic accuracy for disease progression in chronic lymphocytic leukemia. *The Journal of Pathology*, 256(1), 4-14.
3. Kumar, S. A., García-Magariño, I., Nasralla, M. M., & Nazir, S. (2021). Agent-Based Simulators for Empowering Patients in Self-Care Programs Using Mobile Agents with Machine Learning. *Mobile Information Systems*, 2021.
4. Kumar, S. A., Nasralla, M. M., García-Magariño, I., & Kumar, H. (2021). A machine-learning scraping tool for data fusion in the analysis of sentiments about pandemics for supporting business decisions with human-centric AI explanations. *PeerJ Computer Science*, 7, e713.

5. Suryaganesh, M., Arun Samuel, T. S., Ananth Kumar, T., & Navaneetha Velammal, M. (2022). Advanced FET-Based Biosensors—A Detailed Review. *Contemporary Issues in Communication, Cloud and Big Data Analytics*, 273-284.
6. Thiruvikraman, P., Kumar, T. A., Rajmohan, R., & Pavithra, M. (2021). A Survey on Haze Removal Techniques in Satellite Images. *Irish Interdisciplinary Journal of Science & Research (IIJSR)*, 5(2), 01-06.
7. Mostafa, A. M., Kumar, S. A., Meraj, T., Rauf, H. T., Alnuaim, A. A., & Alkhayyal, M. A. (2022). Guava Disease Detection Using Deep Convolutional Neural Networks: A Case Study of Guava Plants. *Applied Sciences*, 12(1), 239.
8. Simsek, E., Badem, H., & Okumus, I. T. (2022). Leukemia Sub-Type Classification by Using Machine Learning Techniques on Gene Expression. In *Proceedings of Sixth International Congress on Information and Communication Technology* (pp. 629-637). Springer, Singapore.
9. Kumar, S. A., Kumar, H., Dutt, V., & Soni, H. (2021, February). Self-Health Analysis with Two Step Histogram based Procedure using Machine Learning. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 794-799). IEEE.
10. Kumar, S. A., Kumar, A., Dutt, V., & Agrawal, R. (2021, February). Multi Model Implementation on General Medicine Prediction with Quantum Neural Networks. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 1391-1395). IEEE.
11. Aof, A. M. B., Awad, E. A., Omer, S. R., Ibraheem, B. A., & Mustafa, Z. A. (2022). A Computer-Aided Diagnoses Program for Leukemia Detection Using Blood Samples. *Journal of Clinical Engineering*, 47(1), 44-49.
12. Kumar, S. A., Kumar, H., Swarna, S. R., & Dutt, V. (2020). Early Diagnosis and Prediction of Recurrent Cancer Occurrence in a Patient Using Machine Learning. *European Journal of Molecular & Clinical Medicine*, 7(7), 6785-6794.
13. Glorindal, G., Mozhiselvi, S. A., Kumar, T. A., Kumaran, K., Katema, P. C., & Kandimba, T. (2021, July). A Simplified Approach for Melanoma Skin Disease Identification. In *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)* (pp. 1-5). IEEE.
14. Kumar, S. A., Kumar, H., Dutt, V., & Dixit, P. (2020). The Role of Machine Learning in COVID-19 in Medical Domain: A Survey. *Journal on Recent Innovation in Cloud Computing, Virtualization & Web Applications* [ISSN: 2581-544X (online)], 4(1).
15. Kumar, K. S., Radhamani, A. S., Sundaresan, S., & Kumar, T. A. (2021). Medical Image Classification and Manifold Disease Identification through Convolutional Neural Networks: A Research Perspective. *Handbook of Deep Learning in Biomedical Engineering and Health Informatics*, 203-225.
16. Kumar, S. A., Kumar, H., Dutt, V., & Swarnkar, H. (2020). COVID-19 Pandemic analysis using SVM Classifier: Machine Learning in Health Domain. *Global Journal on Application of Data Science and Internet of Things* [ISSN: 2581-4370 (online)], 4(1).
17. Suresh, K. K., Sundaresan, S., Nishanth, R., & Ananth, K. T. (2021). Optimization and Deep Learning-Based Content Retrieval, Indexing, and Metric Learning Approach for Medical Images. *Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications*, 79-106.
18. Kumar, S. A., Kumar, H., Dutt, V., & Dixit, P. (2020). Deep Analysis of COVID-19 Pandemic using Machine Learning Techniques. *Global Journal on Innovation, Opportunities and Challenges in Applied Artificial Intelligence and Machine Learning* [ISSN: 2581-5156 (online)], 4(2).
19. Kumar, Tamilarasan Ananth, Rajendrane Rajmohan, Muthu Pavithra, Sunday Adeola Ajagbe, Rania Hodhod, and Tarek Gaber. "Automatic Face Mask Detection System in Public Transportation in Smart Cities Using IoT and Deep Learning." *Electronics* 11, no. 6 (2022): 904.
20. Kumar, S. A., Kumar, H., Dutt, V., & Swarnkar, H. (2020). Role of Machine Learning in Pattern Evaluation of COVID-19 Pandemic: A Study for Attribute Explorations and Correlations Discovery among Variables. *Global Journal on Application of Data Science and Internet of Things* [ISSN: 2581-4370 (online)], 4(2).
21. Das, P. K., Pradhan, A., & Meher, S. (2021). Detection of Acute Lymphoblastic Leukemia Using Machine Learning Techniques. In *Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication* (pp. 425-437). Springer, Singapore.
22. KUMAR, S. A., KUMAR, H., DUTT, V., & SWARNKAR, H. (2019). CONTRIBUTION OF MACHINE LEARNING TECHNIQUES TO DETECT DISEASE IN-PATIENTS: A COMPREHENSIVE ANALYSIS OF CLASSIFICATION TECHNIQUES. *Global Journal on Innovation, Opportunities and Challenges in Applied Artificial Intelligence and Machine Learning* [ISSN: 2581-5156 (online)], 3(1).
23. Pavithra, M., Rajmohan, R., Kumar, T. A., & Sandhya, S. G. (2021). An Overview of Convolutional Neural Network Architecture and Its Variants in Medical Diagnostics of Cancer and Covid-19. *Handbook of Deep Learning in Biomedical Engineering and Health Informatics*, 25-49.

24. Kumar, T. A., Julie, E. G., Robinson, Y. H., & Jaisakthi, S. M. (Eds.). (2021). *Simulation and Analysis of Mathematical Methods in Real-Time Engineering Applications*. John Wiley & Sons.
25. Kumar, A., Chatterjee, J. M., Choudhuri, A., & Rathore, P. S. (2018, November). A Collaborative Method for Minimizing Tampering of Image with Commuted Concept of Frazile Watermarking. In *International Conference On Computational Vision and Bio Inspired Computing* (pp. 985-994). Springer, Cham.
26. KUMAR, A. (2018). FACE RECOGNITION USING HOG-BOW BY INTERNET OF THINGS FOR SECURITY APPLICATIONS. *International Journal of Recent Advances in Signal & Image Processing [ISSN: 2581-477X (online)]*, 2(1).
27. Bhargava, N., Sharma, S., Kumawat, J. R., & Pandey, A. K. (2017, October). An adaptive approach of image fusion (HSI and wavelet approaches) for information refinement in multi image. In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)* (pp. 770-774). IEEE.
28. de Oliveira, J. E. M., & Dantas, D. O. (2021). Classification of Normal versus Leukemic Cells with Data Augmentation and Convolutional Neural Networks. In *VISIGRAPP (4: VISAPP)* (pp. 685-692).
29. Swarna, S. R., Kumar, A., Dixit, P., & Sairam, T. V. M. (2021, February). Parkinson's Disease Prediction using Adaptive Quantum Computing. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 1396-1401). IEEE.
30. Kumar, T. A., Rajakumar, G., & Samuel, T. A. (2021). Analysis of breast cancer using grey level co-occurrence matrix and random forest classifier. *International Journal of Biomedical Engineering and Technology*, 37(2), 176-184.
31. Alam, A., & Anwar, S. (2021). Detecting Acute Lymphoblastic Leukemia Through Microscopic Blood Images Using CNN. *Trends in Wireless Communication and Information Security*, 207-214.
32. Kumar, Abhishek, SwarnAvinash Kumar, Vishal Dutt, Ashutosh Kumar Dubey, and Vicente García-Díaz. "IoT-based ECG monitoring for arrhythmia classification using Coyote Grey Wolf optimization-based deep learning CNN classifier." *Biomedical Signal Processing and Control* 76 (2022): 103638.
33. A.Kumar, S.Kumar, V.Dutt, S.Narang, A.Dubey "A Hybrid Secured Cloud Platform Maintenance based on Improved Attributes. Based Encryption Strategies" published in regular issue in IJIMAI, Indexed by the Science Citation Index Expanded(Web Of Science), Universidad Internacional de La Rioja (UNIR). ISSN 1989-1660.
34. SwarnAvinash Kumar, Harsh Kumar, Vishal Dutt, Himanshu Swarnkar, "Contribution Of Machine Learning Techniques To Detect Disease In Patients : A Comprehensive Analysis Of Classification Techniques" Vol 3 No 1 (2019): *Global Journal on Innovation, Opportunities and Challenges in AAI and Machine Learning*. ISSN 2581-5156.
35. SwarnAvinash Kumar, Kapil Chauhan, Aastha Parihar, "Functionality of Classification and Regression tree in Bioinformatics" Vol 5 No 2 (2021): *Global Journal on Innovation, Opportunities and Challenges in Applied Artificial Intelligence and Machine Learning*. ISSN 2581-5156.
36. Kumar, S.A. (2021), "Corona Recognition Method Based On Visible Light Color Using Artificial Intelligence". AusPat Application No. AU 2021103067(A4).
37. Kumar, S.A. (2021), "An Artificial Intelligence And IoT Based Method For Prevention Of Security Attack On Cloud Medical Data". AusPat Application No. AU 2021102115(A4).
38. Kumar, S.A. (2021), "IOT Based Generic Framework For Computer Security Using Artificial Immune System". AusPat Application No. AU 2021102104(A4).
39. Kumar, S.A. (2021), "IOT Enabled Wall Climbing Robot For Security". AusPat Application No. AU 2021101471(A4).



Impact Factor: 8.118



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details