https://doi.org/10.48047/AFJBS.6.9.2024.5295-5309

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Machine Learning Approaches for Automated Image Analysis in Cell Biology and Structural Biology

Ajay Sudhir Bale[1*], Deepak Pandita[2], Tajinder Kaur [3], Mamta Bahheti[4], Arpita Maitra[5], Sanjeeb Mallick[6]

[1]Dept. of ECE, New Horizon College of Engineering, Bengaluru, India; ajaysudhirbale@gmail.com

[2]Department of Computer Application (MCA), School of Engineering and Technology, Pimpri Chinchwad University, Sate Maval Pune - 412106, Maharashtra (India): pandita.deep@gmail.com

[3]Assistant Professor,Post Graduate Department of Computer Science and Applications ,Sri Guru Teg Bahadur Khalsa College, Sri Anandpur Sahib, India; research.tajinderkaur@gmail.com

[4] Assistant Professor, Department of Computer Science, Hislop College,Civil Lines, Nagpur 440001;

[5] Assistant Professor, Department of Oral medicine and radiology, GNIDSR, Kolkata. PhD scholar, Oral and dental science, JIS UNIVERSITY, Kolkata; arpita6795@gmail.com

[6] Assistant Professor, BBK College, Nagaon, Barpeta; sanjeevmallick@gauhati.ac.in

Corresponding Author: Ajay Sudhir Bale; ajaysudhirbale@gmail.com

Received: 15 March 2024; Revised: 02 April, 2024; Accepted: 28 April, 2024

Volume 6, Issue 9, May 2024

Received: 15 March 2024

Revised: 02 April 2024

Accepted: 28 April 2024

doi: 10.48047/AFJBS.6.9.2024.5295-5309

Abstract: When it comes to studying microscopic picture data in cell biology research, robotic processing of information and machine learning approaches are crucial, as discussed in the paper. Because microscope pictures may be so complicated and varied, it shows how difficult it is to develop suitable algorithms for data processing. Tasks like object identification, movement analysis, and morphometric feature extraction are tackled with the help of machine learning methodologies including supervised learning (classification) and unsupervised learning (clustering). We talk about the possible and actual uses of these methods in structural biology and cell biology, as well as the things to think about and the problems to solve when putting them into practice.

1. Introduction:

Automated data processing is very necessary due to the high data output rate of readily accessible motorised microscopes, which may produce over 105 pictures daily. In addition to making the experimentalist's job easier, computerised statistical analysis guarantees that annotations of big data sets are consistent and objective. Yet, creating appropriate data analysis methods is challenging due to the variety and complexity of microscope photographs. Many studies state that bioimage informatics approaches provide strong answers to certain picture analysis problems such object recognition, movement analysis, and morphometric characteristic measures [1]. The majority of image analysis algorithms, on the other hand, have been tailored to certain biological tests. It is common to need to tweak parameters or possibly re-program the programmes when applying the correct algorithms for various marker or cell types. However, most cell biology labs have significant challenges when it comes to manual computer modifications. This is because of a lack of competence in software engineering and inadequate understanding of the equations beneath image analysis methods. To avoid the need for human intervention in the form of parameter tweaking or established processes, ML attempts to generalise operating rules training from instances . When faced with complicated multi-dimensional statistical duties, like distinguishing shapes that cannot be adequately described by a handful of parameters, machine learning outperforms traditional image processing programmes [2-3]. As a rule, there are two stages of ML. A computer system is trained in its initial stage using a set of data sampling with the purpose of learning from the data's underlying structure and connections. After then, new information from data are fed into the computer system in order to foretell certain characteristics of those samples. According to many studies, the main objective of ML methods is to be able to generalise from small training sets to larger ones and produce correct predictions [4-5].

Classification is one of the most popular areas in ML. By labelling a subset of instances with predetermined categories, the user may construct a training data set using this method. The criteria for class discrimination are automatically inferred by the machine-learning algorithm and may subsequently be applied to the whole dataset. Aiming to infer broad distributional qualities from a small number of annotated instances, this kind of learning is known as "supervised" ML [6-7]. A different kind of ML may glean insights from datasets without any input from humans whatsoever. The objective of "unsupervised" ML is to cluster data points according to a ratio of similarity or to simplify the data in order to make data mining easier. There have been successful applications of unsupervised techniques for phenotypic profiling of pharmacological effects and for exploring unknown phenotypes [8-9]. On the other hand, there are a lot of practical factors to think about and information about the data type and analytical objectives that must be known in order to use ML successfully [10-13]. An effective ML pipeline for the processing of microscopic pictures may be difficult to set up, but this Perspective hopes to help cell biologists do just that. The first thing we'll go over is the process of transforming picture data into units that ML algorithms can use. Next, we'll provide some history of cutting-edge supervised ML techniques and go over some considerations for getting the most out of them. Also covered are the fundamentals of unsupervised ML and its current uses in cellular biology [14-16].

2. Background:

A set of algorithms that can read pictures like a book is based on recent developments in artificial intelligence and machine learning (ML). Applying such deep learning

(DL) algorithms to biological pictures is revolutionising the way imaging data is analysed and understood. New, formerly unattainable tests will be feasible thanks to these developments, which are also poised to make complex analysis commonplace. In [11], the researchers take a look at how DL is influencing cellular image processing and offer a rundown of the mathematical foundations and programming languages used by deep learning, which are important for researchers in the life sciences. The work in [1] take stock of the development in four important areas: enhanced microscopy, tracking of objects, picture segmentation, and classified images [17-18]. The field of biology have been revolutionised by imaging advancements, which have given scientists the ability to study the temporal and spatial shifts that defines systems of life. Microscopes, made possible by advancements in optics, may now capture images at sizes ranging from individual molecules to whole creatures. At the same time, advances in fluorescence sensors have made small-molecule dyes and protein fluorescence brighter, more photostable, and able to cover a wider spectrum [19]. All things considered, these developments pave the way for a wide range of real-time cellular measurements, including but not limited to: long-term imaging of individual molecules; simultaneous measurements of several biosensors; and observations of the evolution of whole organisms. With the advent of spatial genomics, these advancements have also enabled remarkable assessments in fixed samples, with the ability to preserve geographic data while measuring hundreds of mRNA species or dozens of proteins in fixed cells and tissues all at once [20-22].

There has been a rise in the need for analysis of images in the biological sciences that has paralleled these technical developments. In order for contemporary imaging information to be helpful quantifying is becoming more important . Typical tasks involve comparing characteristics among collections of images, classifying images, segmenting images, and tracking objects, such as one cell in a live embryo, across movie frames. Unsupervised image investigation involves contrasting includes of groups of images, classification of images includes predicting a label for an image, and object tracking is tracking an object, such as a stem cell, across multiple frames of a movie. Software archives, application executions, and universal artificial intelligence ecosystems have been built by both academia and corporations to meet this need [23-25].

3. Cell Biology and Structural Biology:

The homeostatic survival of communities of cells depends on their ability to communicate with one another, yet cataloguing all of the ways in which cells do this would be an enormous undertaking. Still, it is possible to have a bird's-eye perspective of these modes, both globally and analytically [26-28], to see where they overlap and where they differ.

To begin, there is the option of non-contact communication, which makes use of an infinite variety of chemicals naturally present in cells' surrounding environment, especially cytokines and growth factors. Afterwards, these elements may have a local effect or are carried to other places by different bodily fluids. This review does not address this communication mode, which is likely least typically studied, because it does not include direct interaction between membranes. However, understanding the practical significance of the distinctions between contact-based and contactless communications is of paramount importance [29].

Then, there is an ever-expanding toolbox of cellular communication mechanisms that all revolve on membrane contacts (Figure 1).

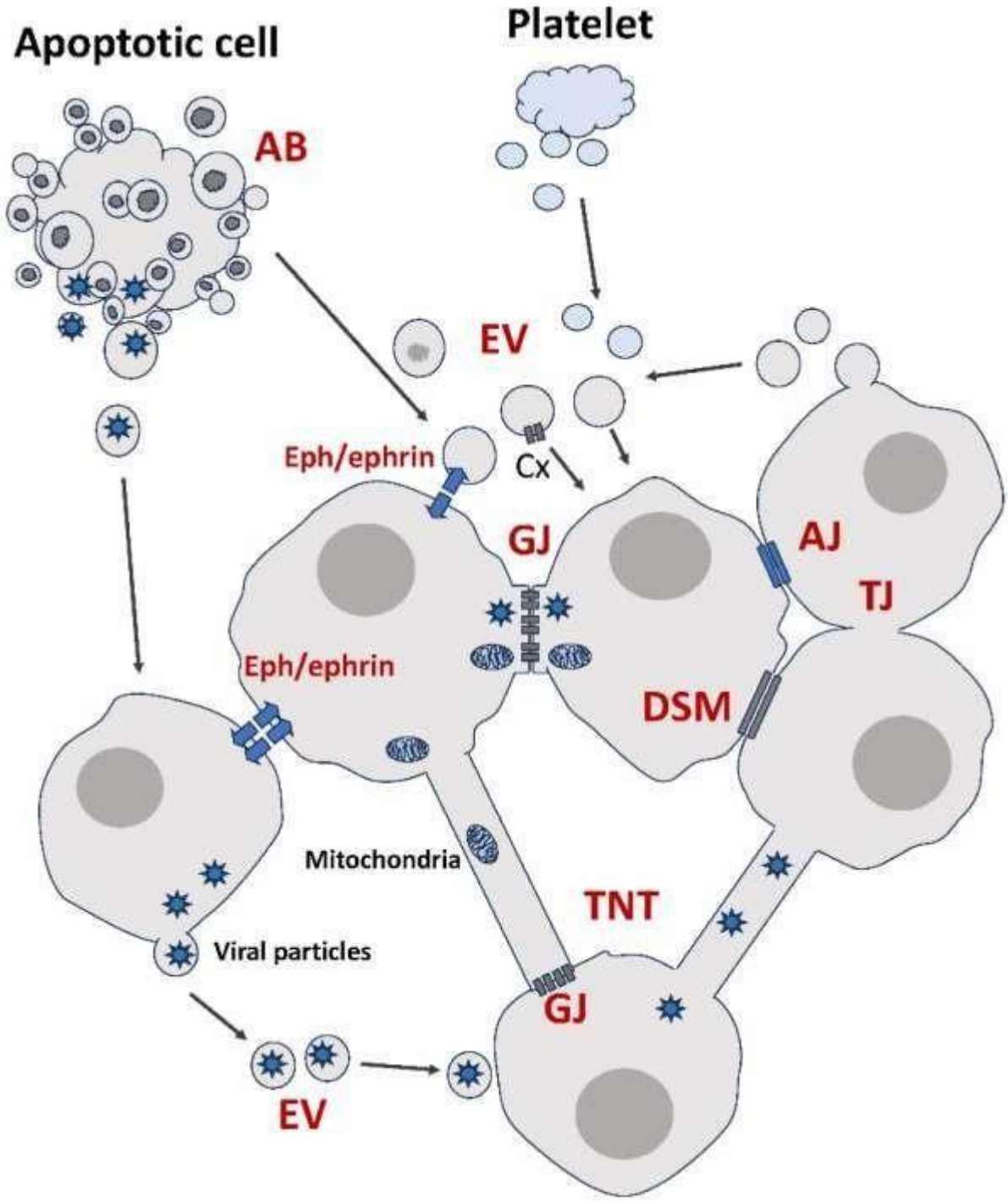

Figure 1: Overview of methods of intercellular interaction centred around membranes [29].

In well-functioning tissues, normal cells are part of a network of interconnected cells that are organised into different tissular compartments, each of which performs a unique role. After that point on, they establish connections via various junctional structures. Desmosomes (DSMs), adherens junctions (AJs), tight junctions (TJs), and gap junctions (GJs) are specialised junctions that enable intercellular interaction as well as cooperation. Macromolecular complexes containing each of these components are structurally and functionally unique. Cell junctions are categorised into three classes according to their functions. Junctions that allow chemicals and electrical impulses to be transmitted are called communicating junctions, or GJs. In epithelial tissues, occluding junctions (TJs) block all molecular transmission between cells.

Finally, focal adhesions (FAs), AJs, and DSMs are structures that are responsible for mechanical adherence between cells or between cells and extracellular matrix (ECM).

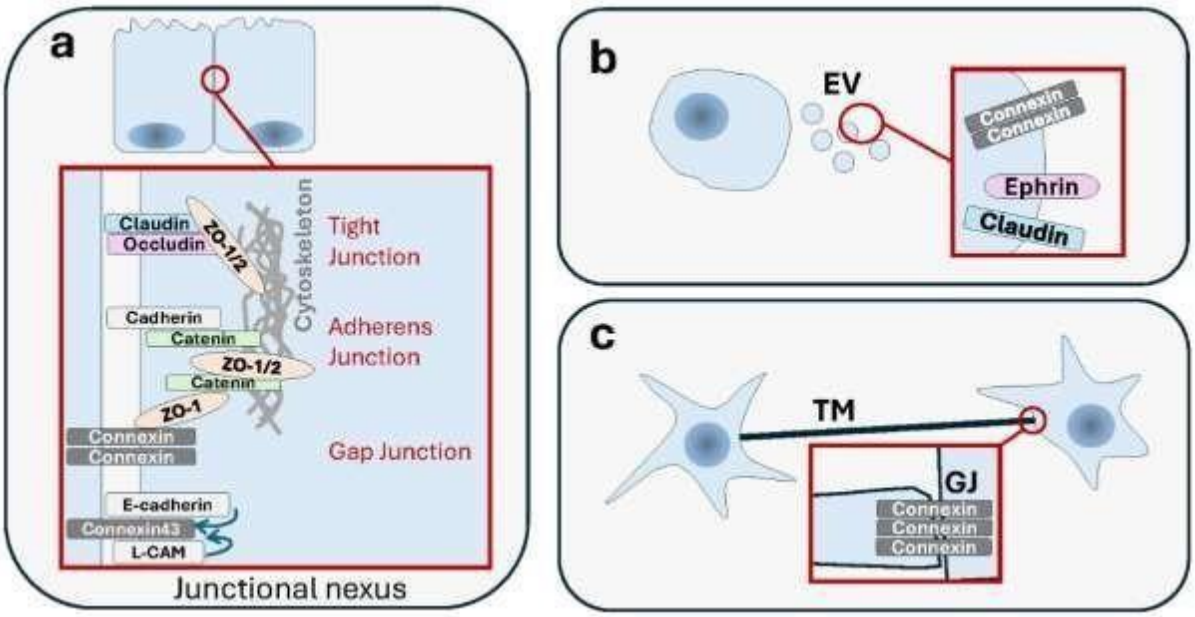

Figure 2: Various forms of communication between cells interact with one another [29]

Investigation of cell-cell cooperation or opposition in different settings, as well as their incorporation into intricate communication webs, is crucial to comprehending the variety of cell-cell interaction modalities. Here, data suggests both structural and functional components are involved in modal interactions as shown in Figure 2 [29].

Worldwide, cancer ranks high among the most prevalent illnesses. There were 1,930,000 confirmed cases of this illness globally in 2022. When compared to other cancers, lung cancer is the worst. The time it takes to diagnose cancer may be drastically reduced with the use of computer-aided diagnostic (CAD). Finding an efficient method to recognise picture characteristics is crucial for using medical pictures as CAD input data. Histopathological diagnoses may be made more accurate with the use of machine learning in computer-aided design systems. When it comes to computer vision (CV), artificial intelligence has made tremendous strides in the last five years. Cancer diagnostics stand to benefit greatly from AI-powered methods for dependency extraction from data. Screening, diagnosis, illness monitoring, therapy planning, and clinical oncology research as a whole are all starting to make extensive use of AI technology. One promising new approach is the creation of computer-driven systems for individualised medicine, which allow for the manufacture of specific medicinal compounds right at the patient's bedside. That is why this strategy may be pivotal to medicine's long-term viability [30]. Radiomics is a branch of artificial intelligence that develops mathematical models for prognosis and prediction via the calculation, proof of identity, and retrieval of picture aspects. For ML-based clinical data analysis, it represents a fresh, emerging, supplemental source for diagnostic data [31]. No intrusive procedures are involved. To further this therapeutic strategy, in silico research need data. A number of parameter correlations may be found. The availability of data for a statistically valid subset of patients is all that is needed. Because of this, the advantages of AI are many.

Concerns about the accuracy of medical diagnostic picture descriptions have an impact on the model used to forecast treatment outcomes and patient care, which is a major issue in modern radiation oncology. One common way this is shown is via the wide range of structural

markings that doctors use to describe patients' photos. Former research by W.C. Sleeman et al. [2] suggested a method to update picture descriptions in accordance with AAPM Task Group 263 (TG- 263) guidelines. Datasets including diagnostic photos for the first 709 lung cancer patients and 752 prostate cancer patients from 40 different hospitals were chosen by the scientists. The VAROQS, or Radiation Oncology Quality Surveillance Programme, was responsible for collecting the training datasets. As supplementary vectors for the comparison study, the authors-built signals of a skeletal framework. Singular value decomposition (SVD) was used to optimise the learning process by lowering the dimensionality of these characteristics. Compared to standard principal component analysis (PCA), SVD delivered an error within 0.1% at a greater speed. For the purpose of verifying the pre-trained models, test datasets were created using data from fifty patients with lung cancer and fifty patients with prostate cancer that were donated by Virginia Commonwealth University (VCU). As for prostate cancer, the RF algorithm showed the greatest performance at 95.06 and lung cancer at 98.77. With the exception of NB, all of the approaches demonstrated 90% accuracy when tested on the prepared data. When tested on the clinical data, the findings were on par with those from the external test dataset at Virginia Commonwealth University (VCU), and MLP emerged as the clear winner. Even when utilising clinical kits, the findings for lung cancer were 95% and for prostate cancer, 91%, since the existence of evidence of a bony structure in all instances boosted the accuracy of the measures. A biomarker is a biological signal that shows how the body responds to treatment or whether biological processes are normal or pathological. Figure 3 is a schematic depicting their study in image analysis, which aims to extract information from a collection of picture data in order to create biomarkers [30].

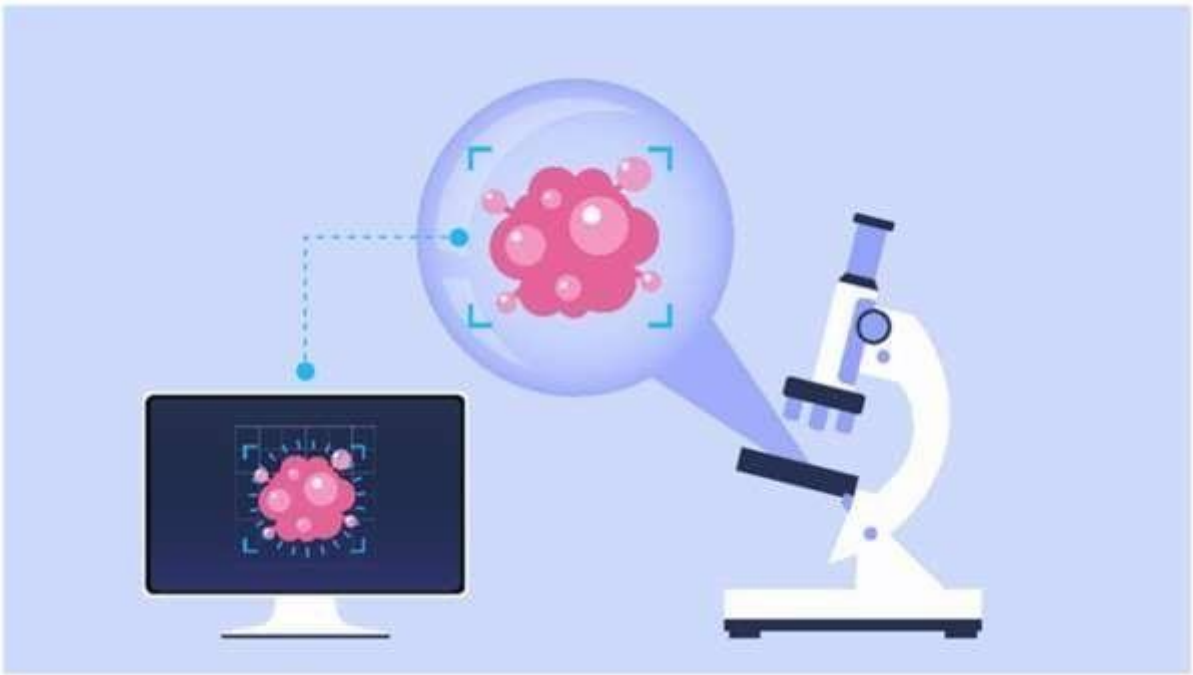

Figure 3: Diagnostics of oncology utilising technology related to computers. [30]

3.1.Long term Health Outlook

Figure 4 shows that there has been a deliberate shift in the creation of cancer guidance systems towards the use of ML algorithms for medical diagnostics and studies, in line with recent technical breakthroughs in the field of computing involving AI.

Figure 4: Long term Health Outlook [30].

With this integration, we can better identify and categorise diseases, draft or revise explanations of diagnostic tests, and predict how a patient will react to treatment. However, there are a number of issues and restrictions with using the equipment in a therapeutic context. Here are the lists that summarise the ML algorithms. They serve a particular purpose in oncological pathology by describing pathologies, assessing the risks of cancer detection, and classifying and predicting disease kinds and stages. The precision, sensitivity, and accuracy of the models used were all diminished in conjunction with the stated data. In each instance, we laid out the practical constraints, and we displayed the statistical results of applying ML algorithms to the training and test datasets, highlighting the test dataset's quantitative and qualitative features [31]. The use of AI to create tailored cancer treatments and the feasibility of building ML-based patient survival prediction models were also evaluated.
When it comes to the prognosis of clear cell renal cell carcinoma, radiomics may be used to build a powerful classifier that reliably predicts patients' overall survival rates [32]. This signature has the potential to help in the detection of individuals at high risk who need more intensive treatment and stricter monitoring plans. With a 95% confidence interval, this model has an area under the curve (AUC) of 0.95-0.98, an accuracy of 0.93-0.98, a specificity of 0.93-0.96, and a sensitivity of 1.0 [32-33].
The preoperative assessment of meningioma consistency is an important component in deciding the extent of surgical intervention and resection. In many studies, research is to use ML classifiers to create a prediction model that can use preoperative medical diagnostic procedures, such as IOUS-E and magnetic resonance imaging (MRI), to determine the characteristics of meningioma consistency. As a starting point, we chose to use data from 18 patients who had surgeries between 2018 and 2020 [33]. The diagnostic procedures of IOUS-E and T 1 WC MRI were used to every instance. Initial steps in analysing the diagnostic data included converting the DICOM electrogram format to hue-saturation-brightness (HSB) and manually segmenting tumours. In order to ascertain the meningioma's consistency, a DT was used as a predictive model; this model was fine-tuned using Range version 3.26, and it was based on the produced groups of intraoperative MTE features that the threshold value was calculated. The six ML algorithms that formed the basis of the classification system were LR, NB, kM, RF, SVM, and MLP [34]. With an accuracy ranging from 61% to 94% and a precision

from 60% to 95% across all algorithms, the models were assessed using the area under the curve (AUC), which yielded values between 0.699 and 0.974. Combining the Information Gain and ReliefF filters with the NB model yielded the greatest results, as seen by an AUC of 0.961 and a 95% accuracy rate, which translates to a 94% classification accuracy.
The capacity to assess the effects of prospective surgical intervention and the hazards of acquiring distant metastases is a major benefit of using ML in data processing. It has the potential to reveal biological details that are inaccessible using standard MRI parameters. In the validation set of patients with soft-tumor sarcoma and distant metastases, it was shown that the AUC could be 0.9510. This indicates that the true positive rate is greater than the false positive rate, suggesting that the mechanism is more likely to score an event as a genuine positive. Gaining insight into the most prevalent organs of distant metastasis may aid in clinical decision making, reducing the financial burden of follow-up exams and the anxiety associated with false positive findings [35].
When it comes to drug discovery, clinical diagnosis, therapy selection and implementation, prognosis, and other clinical concerns, radiomics has shown to be a viable clinical tool. Radiomics is now most often used for determining overall or progression-free survival, predicting therapy sensitivity in tumours, and performing other classification tasks. By creating individualised medication and dosing plans, these methods may one day enhance patient care in AI-based radiation processes [34-35].

3.2.Therapy Routines

Making plans for cancer treatment is no easy task. There are a lot of side effects associated with the individual cancer treatments that are available. Furthermore, cancer treatment aid need a comprehensive strategy and the participation of several experts, including radiologists, surgeons, rehabilitation caretakers, etc. New scientific findings in cancer research also cause shifts in long-held assumptions about how to cure diseases. Figure 5 concludes by classifying the many forms of cancer treatment aid into four categories: initial, curative, supportive, and relapse assistance. Using AI to treat cancer is very challenging due to all of the aforementioned reasons.

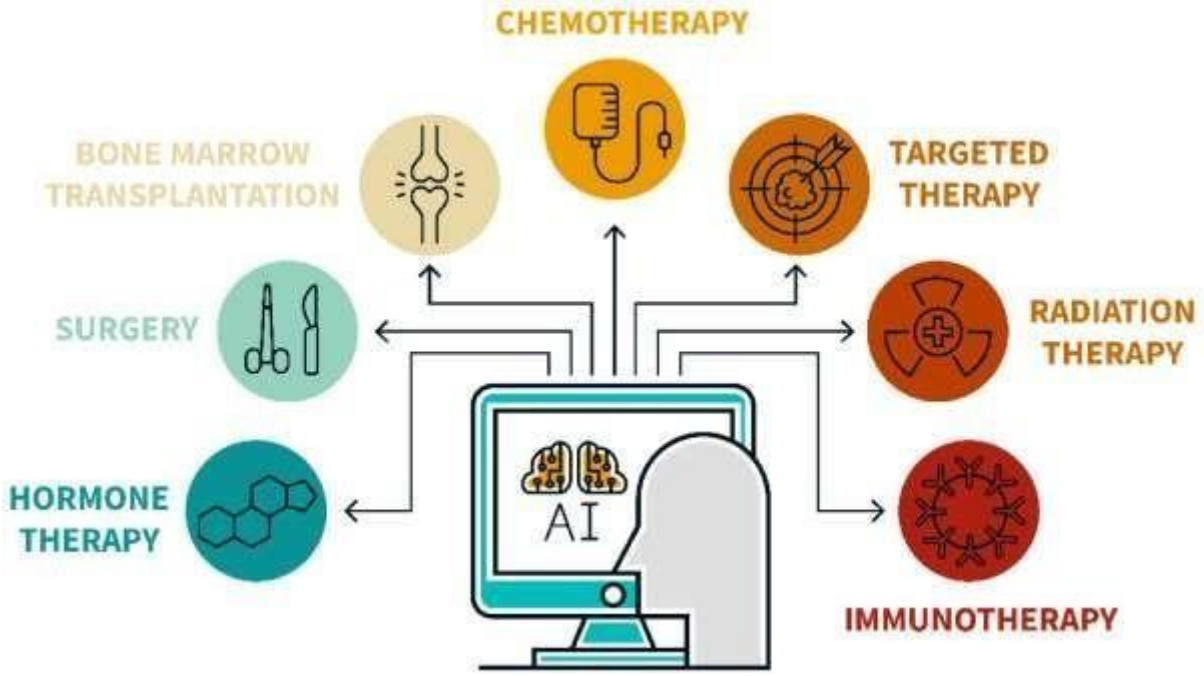

Figure 5: Therapy Routines [30]

Conversely, machine learning (ML) in healthcare is continually demonstrating its value and efficacy. There are many ways in which medical professionals can benefit from ML and Big Data analysis, including keeping up with the ever-increasing amounts of data, cutting costs on resources, improving the quality of treatment and patient care, and making disease diagnoses more quickly. It is challenging to implement such support systems globally due to the need for highly trained computer scientists for such integrations [36].

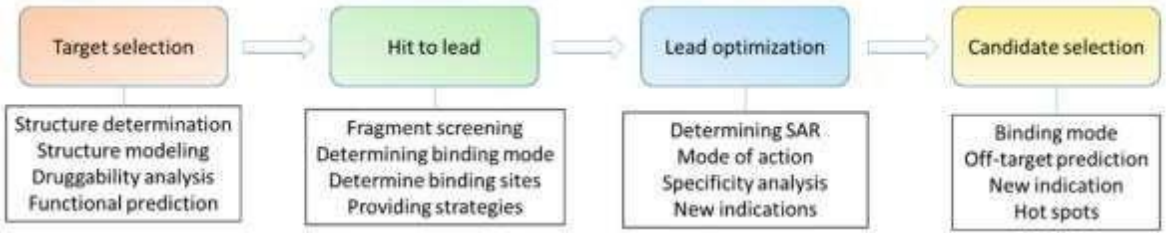

Figure 6: Smaller molecular discovery for drugs and structural biology [36]

In medicine, small molecules play an important role as both therapeutic medications and chemical probes to discover new target functions. As shown in Figure 6, structural biology plays a crucial role in target-based drug development by providing a means to understand how tiny molecules interact with their targets. Building more effective inhibitors and learning more about the roles of small molecules are both made possible by elucidating the structures of targets in complex with their endogenous ligands. When assessing a target's druggability and developing tactics for hit discovery, knowing its structure is crucial. Structured biology will be involved in the hit-to-lead process in three ways: finding hits, validating hits, and optimising compounds. Lead optimisation relies on structural biology to comprehend SAR and forecast specificity. During the preclinical candidate selection process, molecular biology is used to discover the binding modes of the created chemical. These modes explain their SAR and can additionally foresee resistance to drugs and off-target consequences (Figure 6). Compound screening with computational assistance and design of drugs will also be crucial in drug development if a protein structure is accessible [36].

4. Analysis of images and the associated challenges:

In order to quantify minor traits and decrease subjective bias, image analysis software is currently employed across biomedical research. This software is specifically designed to operate with microscope pictures. Another tool that is revolutionising contemporary science is the automated microscope. Multidimensional imaging (time-lapse and three-dimensional [3D]) also generates massive data sets that need automated processing, and experiments evaluating chemical compounds or genetic perturbations may grow to millions of perturbations. Algorithms run by computers must accurately identify cells, subcompartments, or organisms in this mountain of data, and then extract the metrics that describe them. A number of types of biologist-focused image analysis software are vying for dominance in the market for automated microscopy. These include standalone commercial image processing tools like Imaris—Bitplane and MetaMorph—Molecular Devices and Elements—Nikon, as well as free open-source packages like ImageJ/Fiji, CellProfiler, Icy, and KNIME. When used with a microscope, commercial software is often the most user-friendly option. There is an emphasis on practicality, especially for pharmaceutical-related applications, despite the fact that price and rigidity may restrict uptake. But researchers still can't see how commercial software is analysing their data or change an algorithm's approach if they want to since the

code is proprietary [37]. Here, in [38], the authors show that supervised machine-learning models can be trained to distinguish between hyperosmotic-stressed yeast cells and healthy cells by combining support vector machine and random forest techniques. Digital micrographs of nuclear areas of interest from 2000 Saccharomyces cerevisiae cells were analysed using fractal, grey level co-occurrence matrix (GLCM), and discrete wavelet transform methods. Of these cells, 1000 were subjected to hyperosmotic conditions while 1000 served as controls. For deciding if a cell was part of the test or control group, the support vector machine had a respectable classification accuracy of 71.7%. The success rate for classification of the random forest model was 79.8%, which was higher than that of the support vector machine. In order to forecast different physiological and pathological events related to osmotic stress, these results may be used as a foundation for building AI-based algorithms that employ GLCM, fractal, and wavelet data to categorise healthy and injured cells. The research demonstrates that yeast cells subjected to hyperosmotic stress undergo notable alterations in nuclear fractal dimension and nuclear texture, as measured by GLCM and DWT techniques. Additionally, we suggest feeding GLCM, fractal, and DWT indicators into ML models like random forests and support vector machines. Although trained on a limited sample, the models were able to discriminate between healthy and treated cells with a respectable degree of discriminatory power and classification accuracy. To further develop AI-based approaches for identifying osmotic shock damage in cells and their components, these discoveries and models may be used as a basis. [39-40] Computing approaches for analysing cell nuclei texture using tools like GLCM, DWT, and fractal analysis is an emerging and under-tested area of cell biology. In terms of their validity and sensitivity, these approaches have a number of drawbacks [41-43]. The structural homogeneity in cell and tissue micrographs may be uncovered using these methods, which can be applied to a variety of cell populations in vitro and in vivo. It ought to be mentioned, too, that histologic uniformity is not always consistent with textural uniformity. Additional verification from future research is also necessary for the use of DWT indicators to assess textural heterogeneity. Just as our work uses fractal analysis—a mathematics and biological method—to infer the complexity of a signal, subsequent studies in this area will need to validate its various possible uses. The signal in question might be one-dimensional or two-dimensional [44].

Developing machine learning models to evaluate cell damage using Saccharomyces cerevisiae is nothing new. Angular second moment and inverse difference are two GLCM indicators that were recently studied to determine the effect of sublethal dosages of ethanol. Less textural regularity and local homogeneity of cell nuclei were seen in response to ethanol, which is comparable to the effects of hyperosmotic stress brought on by NaCl. We suggested GLCM analysis as well as multilayer perceptron, binomial logistic regression, and random tree machine learning models [44-45]. The neural network outperformed the other two models with the best area under the receiver operating characteristic curve (AUC=0.87), but all three models had good classification accuracy. Since alcohol may induce hyperosmotic stress in some contexts, the alterations in GLCM markers are partly consistent with our present study's findings. On the other hand, it's worth mentioning that these changes in the nuclear structure are probably caused by ethanol damaging the cells' DNA or causing chromatin to reorganise in nuclei as a consequence of certain signalling pathways linked to ethanol damage. Micrographs of cells and tissues may have their structural complexity and degree of detail indirectly measured using fractal analysis. Since fractal dimension may help evaluate the branching patterns of axons and dendrites, its most widespread application in microscopy is likely to be in the study of neurons in the central and peripheral nervous systems. The fractal dimension and other fractal markers in chromatin and cell nuclei have been the subject of several attempts at quantification. Heterochromatin seems to have a different fractal dimension than

euchromatin. An alternative to the traditional equilibrium model of chromatin organisation, the so-called "fractal globule" model has been proposed in recent years. We still don't know much about the fractality of chromatin and DNA, two macromolecules that share certain self-similarity features. Our study's data for training the ML model was generated entirely from fractal analysis, which we used to detect minute morphological changes in cell nuclei.

Cell volume decreases significantly and many adaption mechanisms are activated when Saccharomyces cerevisiae yeast cells are exposed to a hyperosmotic environment. The viability is mostly maintained, and the cells include many genes that are involved in osmo protection; moreover, these cells are very resistant to high tonicity. Severe osmotic stress in yeast frequently leads to cell cycle halt, growth inhibition, and decreased metabolic activity. To compensate for the high osmolarity of the extracellular space, sodium chloride-compatible solutes such as glycerol, trehalose, and erythritol are being manufactured. Furthermore, several stress-response pathways are activated by hyperosmotic stress brought on by NaCl or other osmotically active substances. These pathways include the stress-activated protein kinase pathway (SAPK) and the high osmolarity glycerol (HOG) route [46-49]. As a result, nuclear chromatin patterns may rearrange themselves, and DNA transcription and gene expression undergo substantial modifications. In addition, hyperosmotic stress could lead to an increase in ROS generation and oxidative damage to DNA in cells. It is reasonable to assume that the epigenetic changes that occur as a cell adapts to damage are the main factor contributing to the observed changes in computational texture indicators, although it is not yet clear which of these processes is responsible for them. Perhaps the most intriguing part of our work is how computational approaches picked up on minute morphological changes in cells that were invisible under conventional microscopes. There were no obvious symptoms of necrosis, significant nuclear damage, or programmed cell death in the cells that were examined. It was subjectively impossible to distinguish between the two sets of cells, even for the researcher with extensive background in microscopy and cell biology, since their morphologies were so similar. The experimental group was not the only one that experienced small changes in nuclear structure. For example, the darkened areas on the nuclear periphery (Figure 1) could have been caused by a physiological variation specific to this cell type or by a variation in the amount of light and white balance used during microscopy and cell acquisition. Computing techniques, however, revealed substantial group variations in fractal, textural, and wavelet indicators. This may increase the scientific utility of these approaches in disease by demonstrating their ability to identify distinct morphological features. These techniques might one day be a component of pathology-focused computer-aided diagnostic systems that help with things like identifying different kinds of diseased cells or distinguishing between damaged and healthy cells in a range of lab and clinical settings. But before that can happen, we need to fix the many stages and other problems with these techniques. Our work has several major caveats that could reduce its usefulness for future research in microscopy and cell biology. In this field of study, all of the computational approaches that have been deployed are still in their infancy. Their validity, accuracy, and overall quality assurance have not been thoroughly tested. For the majority of cell types and tissues, the reliability of the approaches between and among observers is unknown. This is still a big problem for modern histology and pathology, even if there have been attempts to provide quality assurance in the past. The second constraint is the large variation in fractal and GLCM indicator values between various substrates and experimental conditions, which is a major drawback. For instance, when using a different image acquisition equipment or producing micrographs of varying sizes and resolutions might lead to significantly different values for the angular second moment and inverse difference moment. Everything from the kind of microscope to the imaging equipment to the default options in imaging software may have a significant impact on different microscope settings including exposure, hue, saturation, and white balance. Lastly, it's important to note that just because we

saw changes in computational indicators in one specific yeast culture doesn't mean that other cultures will also exhibit same changes. one is especially true when different fixation and staining techniques are used. Future research should use yeast-specific staining protocols to visualise chromatin structure in order to make firm conclusions on the alterations in chromatin textural patterns under hyperosmotic stress and the possible utility of pattern recognition methods in this domain. Five GLCM indicators—angular second moment, inverse difference moment, correlation, and variance—were the only ones we measured in our study. The number of textural characteristics that may be obtained is much larger when using GLCM and comparable methods of textural analysis. For instance, there are several quantifications such as entropy, sum entropy, difference entropy, information measures of correlation, maximum correlation coefficient, and others. Before incorporating this strategy into modern cell biology and pathology practice, future research should use all available characteristics to construct the most effective ML models. Fractal analysis follows the same line of thinking as it allows for the quantification of non-fractal characteristics as well. To better understand the variations in complexity, it is common practice to compute the fractal dimension in conjunction with the lacunarity feature, which is a measure of the amount of "gappiness" inside a fractal. This is the most relevant example. It would be fascinating to see how RF and SVM models trained with GLCM and lacunarity data perform in terms of discriminatory power and classification accuracy down the road. It is important to note that our study's use of a machine learning technique is not without its limits. Training supervised ML algorithms involves providing examples of input and (correct) output or goal data, such as support vector machine and random forest. Neural networks, non-random forest decision trees, and binomial logistic regression models are among the other models that might be useful. When trained with the GLCM and DWT data, some of these methods could be more effective at seeing patterns and producing more accurate classifications. It would be wise to create and evaluate all potential ML models in the future, then choose the best one to build an application for the web or another platform. Training and testing samples were small in our research. Perhaps a more comprehensive model (like the one based on a multilayer perceptron network) might be trained with more data, particularly GLCM data, and achieve remarkable results in cell categorization. Finally, ML has a major drawback when it comes to interpretability, which is a problem for most ML models. Even with random forests and support vector machines, understanding how the model works and the underlying principles that influence its cell categorization choices remains a challenge, even if we may get intriguing performance results. Because of this and the basic problem of not ensuring the quality of computational methods used to generate data, the findings are not very reproducible. These problems need to be addressed in future studies before these models may be employed in modern pathology and other disciplines' diagnostic procedures and research.

Discussions and Conclusions:

Automated picture analysis in structural and cell biology might benefit greatly from machine learning approaches, yet these methods aren't without their limits and problems. Additional validation and quality assurance are necessary for computational approaches such as fractal analysis, discrete wavelet transform (DWT), and grey-level co-occurrence matrix (GLCM) to guarantee accurate interpretation of data from microscopic images. To successfully include ML models into diagnostic processes and research applications, problems with data quality, repeatability, and interpretability must be addressed. It is also critical to deal with the fact that computational indicator values could vary depending on the experimental circumstances and the picture capture equipment. We need more in-depth analyses of machine learning models in

the future that take into account a variety of textural features and investigate other methods, such as neural networks. In order to realise machine learning's revolutionary promise in image analysis and to propel new discoveries in cell biology and structural biology, these obstacles must be overcome.

References:

1. Moen, E., Bannon, D., Kudo, T. et al. Deep learning for cellular image analysis. Nat Methods 16, 1233–1246 (2019). https://doi.org/10.1038/s41592-019-0403-1
2. Sommer, Christoph, and Daniel W. Gerlich. "Machine learning in cell biology–teaching computers to recognize phenotypes." Journal of cell science 126.24 (2013): 5529-5539.
3. Rittscher, Jens. "Characterization of biological processes through automated image analysis." Annual Review of Biomedical Engineering 12 (2010): 315-344.
4. Kousaka, Jin, Atsuko H. Iwane, and Yuichi Togashi. "Automated Cell Structure Extraction for 3D Electron Microscopy by Deep Learning." arXiv preprint arXiv:2405.06303 (2024).
5. Jan, Mahta, et al. "From pixels to insights: Machine learning and deep learning for bioimage analysis." BioEssays 46.2 (2024): 2300114.
6. Schmitt, Matthew S., et al. "Machine learning interpretable models of cell mechanics from protein images." Cell 187.2 (2024): 481-494.
7. Goshisht, Manoj Kumar. "Machine learning and deep learning in synthetic biology: Key architectures, applications, and challenges." ACS omega 9.9 (2024): 9921-9945.
8. Brun, Aurélien, et al. "A new bio imagery user-friendly tool for automatic morphometry measurement on muscle cell cultures and histological sections." Scientific Reports 14.1 (2024): 3108.
9. Bender, Steen WB, et al. "SEMORE: SEgmentation and MORphological fingErprinting by machine learning automates super-resolution data analysis." Nature Communications 15.1 (2024): 1763.
10. Wang, Xiaoli, et al. "Deep learning-based image analysis for in situ microscopic imaging of cell culture process." Engineering Applications of Artificial Intelligence 129 (2024): 107621.
11. Carreras-Puigvert, J., & Spjuth, O. (2024). Artificial intelligence for high content imaging in drug discovery. Current Opinion in Structural Biology, 87, 102842.
12. Shaban, Muhammad, et al. "MAPS: Pathologist-level cell type annotation from tissue images through machine learning." Nature Communications 15.1 (2024): 28.
13. Nogales, Eva, and Julia Mahamid. "Bridging structural and cell biology with cryo-electron microscopy." Nature 628.8006 (2024): 47-56.
14. Kulkarni, Tanmay, et al. "Machine learning-based approach for automated classification of cell and extracellular matrix using nanomechanical properties." Materials Today Bio 25 (2024): 100970.
15. Li, Hui, et al. "Machine learning-assisted high-content imaging analysis of 3D MCF7 microtissues for estrogenic effect prediction." Scientific Reports 14.1 (2024): 2999.
16. Nelson, Anders R., et al. "Logic-based mechanistic machine learning on high-content images reveals how drugs differentially regulate cardiac fibroblasts." Proceedings of the National Academy of Sciences 121.5 (2024): e2303513121.
17. Guo, Zhiye, et al. "Diffusion models in bioinformatics and computational biology." Nature reviews bioengineering 2.2 (2024): 136-154.
18. Duran, Imanol, et al. "Detection of senescence using machine learning algorithms based on nuclear features." Nature Communications 15.1 (2024): 1041.
19. Thiermann, Ryan, et al. "Tools and methods for high-throughput single-cell imaging with the mother machine." Elife 12 (2024): RP88463.
20. Wiggin, Noah, et al. "Empowering high-throughput high-content analysis of 3D tumor models: Open-source software for automated non-confocal image analysis." bioRxiv (2024): 2024-02.

21. Dunster, Gideon, Matheus Palhares Viana, and Susanne M. Rafelski. "Streamlining deep-learning-based segmentation methods for microscopy images." Biophysical Journal 123.3 (2024): 430a-431a.
22. Egebjerg, Jacob Marcus, et al. "Automated quantification of vacuole fusion and lipophagy in Saccharomyces cerevisiae from fluorescence and cryo-soft X-ray microscopy data using deep learning." Autophagy 20.4 (2024): 902-922.
23. Kim, Suryong, et al. "Angio-Net: deep learning-based label-free detection and morphometric analysis of in vitro angiogenesis." Lab on a Chip (2024).
24. Seal, Srijit, et al. "From pixels to phenotypes: Integrating image-based profiling with cell health data as BioMorph features improves interpretability." Molecular Biology of the Cell 35.3 (2024): mr2.
25. Müller, Andreas, et al. "Modular segmentation, spatial analysis and visualization of volume electron microscopy datasets." Nature Protocols (2024): 1-31.
26. Pham, Dan L., et al. "Perspectives on label-free microscopy of heterogeneous and dynamic biological systems." Journal of Biomedical Optics 29.S2 (2025): S22702-S22702.
27. Moshkov, Nikita, et al. "Learning representations for image-based profiling of perturbations." Nature Communications 15.1 (2024): 1594.
28. Zhu, Hongkang, et al. "A machine learning approach for systematic quantification of in vivo cell geometries during embryogenesis." Biophysical Journal 123.3 (2024): 550a-551a.
29. Ebrahim, T.; Ebrahim, A.S.; Kandouz, M. Diversity of Intercellular Communication Modes: A Cancer Biology Perspective. Cells 2024, 13, 495. https://doi.org/10.3390/cells13060495
30. Varlamova, E.V.; Butakova, M.A.; Semyonova, V.V.; Soldatov, S.A.; Poltavskiy, A.V.; Kit, O.I.; Soldatov, A.V. Machine Learning Meets Cancer. Cancers 2024, 16, 1100. https://doi.org/10.3390/cancers16061100
31. Bonlawar, J., Setia, A., Challa, R.R., Vallamkonda, B., Mehata, A.K., Vaishali, , Viswanadh, M.K., Muthu, M.S. (2024). Targeted Nanotheransotics: Integration of Preclinical MRI and CT in the Molecular Imaging and Therapy of Advanced Diseases. Nanotheranostics, 8(3), 401-426. https://doi.org/10.7150/ntno.95791.
32. Pasala, P. K., Rudrapal, M., Challa, R. R., Ahmad, S. F., Vallamkonda, B., & R., R. B. (2024). Anti-Parkinson potential of hesperetin nanoparticles: in vivo and in silico investigations. Natural Product Research, 1–10. https://doi.org/10.1080/14786419.2024.2344740
33. Suseela, M. N. L., Mehata, A. K., Vallamkonda, B., Gokul, P., Pradhan, A., Pandey, J., ... & Muthu, M. S. (2024). Comparative Evaluation of Liquid-Liquid Extraction and Nanosorbent Extraction for HPLC-PDA Analysis of Cabazitaxel from Rat Plasma. Journal of Pharmaceutical and Biomedical Analysis, 116149. https://doi.org/10.1016/j.jpba.2024.116149
34. Chakravarthy, P.S.A., Popli, P., Challa, R.R. et al. Bile salts: unlocking the potential as bio-surfactant for enhanced drug absorption. J Nanopart Res 26, 76 (2024). https://doi.org/10.1007/s11051-024-05985-6
35. Setia, A., Vallamkonda, B., Challa, R.R., Mehata, A.K., Badgujar, P., Muthu, M.S. (2024). Herbal Theranostics: Controlled, Targeted Delivery and Imaging of Herbal Molecules. Nanotheranostics, 8(3), 344-379. https://doi.org/10.7150/ntno.94987.
36. Dhamija P, Mehata AK, Tamang R, Bonlawar J, Vaishali, Malik AK, Setia A, Kumar S, Challa RR, Koch B, Muthu MS. Redox-Sensitive Poly(lactic-co-glycolic acid) Nanoparticles of Palbociclib: Development, Ultrasound/Photoacoustic Imaging, and Smart Breast Cancer Therapy. Mol Pharm. 2024 May 5. doi: 10.1021/acs.molpharmaceut.3c01086. Epub ahead of print. PMID: 38706253.
37. Eranti, Bhargav and Mohammed, Nawaz and Singh, Udit Narayan and Peraman, Ramalingam and Challa, Ranadheer Reddy and Vallamkonda, Bhaskar and Ahmad, Sheikh F. and DSNBK, Prasanth and Pasala, Praveen Kumar and Rudrapal, Mithun, A Central Composite Design-Based Targeted Quercetin Nanoliposomal Formulation: Optimization and Cytotoxic Studies on

MCF-7 Breast Cancer Cell Lines. Available at SSRN: https://ssrn.com/abstract=4840349 or http://dx.doi.org/10.2139/ssrn.4840349

38. Setia A, Challa RR, Vallamkonda B, Satti P, Mehata AK, Priya V, Kumar S, Muthu MS. Nanomedicine And Nanotheranostics: Special Focus on Imaging of Anticancer Drugs Induced Cardiac Toxicity. Nanotheranostics 2024; 8(4):473-496. doi:10.7150/ntno.96846. https://www.ntno.org/v08p0473.htm
39. Bale, A. S. ., Tiwari, S. ., Khatokar, A. ., N, V. ., & Mohan M S, K. . (2021). Bio-Inspired Computing-A Dive into Critical Problems, Potential Architecture and Techniques. Trends in Sciences, 18(23), 703. https://doi.org/10.48048/tis.2021.703
40. A. S. Bale et al., "Nanosciences fostering cross domain engineering applications," Mater. Today, vol. 43, pp. 3428–3431, 2021.
41. J. Aditya Khatokar et al., "Carbon nanodots: Chemiluminescence, fluorescence and photoluminescence properties," Mater. Today, vol. 43, pp. 3928–3931, 2021.
42. N. Vinay, et al., "Artificial intelligence as a tool for conservation and efficient utilization of renewable resource," Artificial Intelligence for Renewable Energy Systems. Wiley, pp. 37–77, 18-Feb-2022.
43. Ajay Sudhir Bale, S. Saravana Kumar, S. Varun Yogi, Swetha Vura, R. Baby Chithra, N. Vinay , P. Pravesh, "Network and security leveraging IoT and image processing: A quantum leap forward," Sciencedirect.com, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/B9780323902403000084. [Accessed: 14-Jun-2024].
44. A. S. Bale et al., "Nanotechnology as a tool for treating cancerous tumors," Mater. Today, vol. 43, pp. 3847–3851, 2021.
45. S. Joy, R. et al., "A Comparative Study on Recent Trends in Iris Recognition Techniques," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1521-1525, doi: 10.1109/ICEARS53579.2022.9752355.
46. A. S. Bale et al., "5G wireless communication and its adverse effects on the human body: Distinguishing falsehoods from reality," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 5s, pp. 101–112, 2024.
47. M. B. Savadatti et al., "Keyboard acoustic side channel attacks on android phones in the AI era of digital banking: An elementary review," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 4s, pp. 292–300, 2024.
48. M. B. Savadatti et al., "Machine learning based Tesla stock analysis: A theosophical review," in 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2023, pp. 1–6.
49. M. B. Savadatti et al., "Machine learning applications for 3D-printed polymers and their composites," in Advances in Chemical and Materials Engineering, Hershey, PA: IGI Global, 2023, pp. 239–260.