IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11794

A Study on Diagnosis of Breast Cancer using Machine Learning

D Guna Karthikeya[1], Keerthi Teja N[2], Rishidevrath Shetty[3], Supreeth M[4], Nivedh A[5]

Department of Artificial Intelligence and Machine Learning, Dayananda Sagar Academy of Technology and Management[1-5]

Abstract: Breast cancer is a prevalent and life-threatening disease affecting millions of women globally. Early and accurate diagnosis is crucial for successful treatment and improved patient outcomes. In recent years, machine learning has emerged as a powerful tool in the field of medical imaging and diagnostics, offering potential advancements in breast cancer detection and classification.

Several machine learning algorithms, including support vector machines (SVM), artificial neural networks (ANN), random forests, and convolutional neural networks (CNN), are employed to build predictive models based on the extracted features. The models are trained and evaluated using a comprehensive dataset comprising a diverse range of breast images, annotated by experienced radiologists.

Keywords: Breast Cancer, convolutional neural networks (CNNs)

I. INTRODUCTION

Breast cancer remains a critical global health challenge, representing one of the most prevalent and life- threatening forms of cancer among women. Early and accurate diagnosis is fundamental for timely intervention, improved treatment outcomes, and ultimately, saving lives. Over the past decade, the integration of machine learning into the realm of medical diagnostics has shown immense potential for enhancing the accuracy, efficiency, and objectivity of breast cancer detection and classification.

This literature review focuses on the application of machine learning techniques for the diagnosis of breast cancer, utilizing diverse datasets that encompass medical imaging, genetic markers, clinical history, and other relevant information. Machine learning algorithms, by their ability to learn patterns and make predictions from complex data, offer an avenue for automated and data- driven analysis that can aid clinicians and radiologists in their diagnostic decisions.

The utilization of machine learning in breast cancer diagnosis involves several key steps, including feature extraction, feature selection, model development, and evaluation. Feature illuminate the transformative role that gait analysis can assume in fortifying security surveillance across diverse environments, encompassing settings ranging from airports to smart buildings and beyond. As we delve deeper into the convergence of biometrics and security, gait analysis emerges as a groundbreaking approach with the ability to reshape our perception and utilization of security technologies. This examination aims to shed light on the transformative role that gait analysis can play in fortifying security surveillance across diverse. Various machine learning algorithms, ranging from classical approaches like support vector machines and decision trees to more advanced techniques such as artificial neural networks and deep learning models like convolutional neural networks (CNNs), have been employed for the development of predictive models. These models can classify breast lesions into malignant or benign categories, providing crucial guidance to clinicians. environments, ranging from airports to smart buildings and beyond.

This literature review aims to synthesize and critically analyse the existing body of research, discussing the methodologies, datasets, performance metrics, and challenges associated with utilizing machine learning in breast cancer diagnosis. By summarizing the current state-of-the-art, identifying gaps, and suggesting potential future research directions, this review aims to provide a comprehensive understanding of the role and potential of machine learning in advancing breast cancer diagnosis, ultimately contributing to improved patient care and outcomes.

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11794

II. PREVIOUS WORK

Automating surveillance systems has become imperative to reduce human errors, combat crimes effectively, and thwart potential terrorist threats. The integration of biometric technologies into closed-circuit television (CCTV) represents a significant leap in automating the process of tracking criminals. Unlike other biometric methods, which may encounter challenges in surveillance scenarios, an individual's walking pattern can be reliably captured and recognized from a distance, even with low-resolution video.

The primary concern centres around the safety and security of a nation's citizens, especially in light of the alarming increase in crime rates and unexpected attacks. CCTV systems rely on remotely positioned cameras to transmit and store real-time video streams. Since the introduction of the gait energy image (GEI) by Han, numerous refined approaches have emerged. Various methods, as put forth by researchers such as Lain, Theekhanont, Sivapalan, and Xue, leverage GEI for gait recognition and have showcased its effectiveness.

This paper presents a novel approach that capitalizes on the Gabor magnitude of GEI and a dimension-reduction technique to extract distinctive features. To achieve an exceptionally high level of security and accuracy, surpassing conventional methods, the adoption of multimodal biometrics becomes imperative. This approach combines multiple biometric recognition technologies to establish an individual's identity. Feature extraction techniques are broadly categorized into two fundamental approaches: selecting individual features and characterizing geometrical relationships. Holistic methods, including principal component analysis (PCA), linear discriminate analysis (LDA), and independent component analysis (ICA), extract appearance information from the entire image while reducing dimensionality.These extracted features significantly enhance classification performance by eliminating irrelevant data.

Previous research has proposed an array of techniques for gait analysis, ranging from trajectory and velocity analysis to discrete symmetric operators and continuous hidden Markov models (HMMs). Han and Bhanu's work extensively utilized the gait energy image (GEI) and incorporated a statistical feature extraction approach, alongside fusion strategies to enhance recognition.

In another approach, Eigen space transformation via Principal Component Analysis (PCA) was deployed to reduce feature space dimensionality, followed by supervised pattern classification techniques within the lower-dimensional Eigen space for recognition. Su and Zanga employed fuzzy principal component analysis for recognition.

In conclusion, this paper strives to make a significant contributionto the field of gait analysis and biometric recognition by introducing a pioneering method. It also underscores the critical importance of automating surveillance systems to bolster securityand ensure public safety.

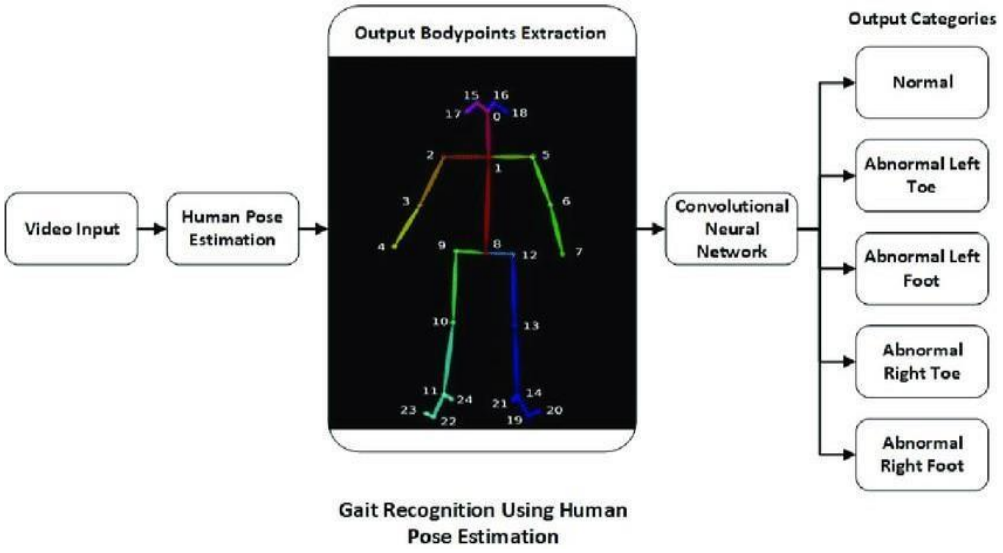

Fig 2: Gait Recognition Using Human Pose

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11794

III. PROPOSED METHODOLOGY

We are going to deliberate about the different methodology that can be used in the gait analysis of security surveillance system.

[1]. SEGMENTATION
[2]. EROSION
[3]. DILATION
[4]. SILHOUETTE
[5]. FEATURE EXTRACTION

3.1 SEGMENTATION

3.1.1 Edge Based Segmentation Method

Edge detection techniques are a well-established field within image processing, playing a pivotal role in image analysis by identifying abrupt changes in intensity. These changes often correspond to object boundaries or significant features in an image, which are challenging to discern using traditional intensity values alone.

Edge detection methods primarily identify edges where either the first derivative of intensity surpasses a predefined threshold or where the second derivative exhibits zero crossings. In edge- based segmentation approaches, the primary objective is to first detect these edges and then connect them to outline object boundaries, enabling the segmentation of specific regions within an image.

Two fundamental categories of edge-based segmentation methods include Gray histograms and Gradient-based methods. These techniques aim to detect edges through the application of well- known edge detection operators such as the Sobel operator, Canny operator, Robert's operator, among others. Typically, the outcome of these methods is a binary image, where pixels are assigned values to indicate the presence of edges

3.1.2 Region-Based Segmentation Method

Region-based segmentation methods are strategies that divide an image into multiple regions with similar characteristics. This approach relies on two fundamental techniques to achieve segmentation**.**

3.1.3 Region growing methods.

Region growing-based segmentation methods are techniques used to divide an image into various regions based on the expansion of seeds, which are initial pixels. These seeds can be chosen manually, relying on prior knowledge, or automatically, depending on the specific application. The growth of seeds is then controlled by the connectivity between pixels, often guided by prior knowledge of the subject matter, and can be terminated accordingly**.**

2. Estimation of an approximate background from a sequence of images depicting people walking. The primary mean image is computed by averaging the gray-level values at each pixel over the entire image sequence (illustrated in Fig.1 (b)). Let I_k(x, y), where k=1, 2, ..., N, represent the sequence of N images. Background images, denoted as b(x, y), can be calculated as follows: b(x, y) = median(I_k(x, y)), where k=1,
2, ..., N.

3.2 EROSION

Erosion is one of the two fundamental operators in mathematical morphology, with the other being dilation. While erosion is typically applied to binary images, certain variations can also operate on grayscale images. In its basic form, this operator acts on a binary image, and its primary effect is to gradually erode or diminish the boundaries of regions

IARJSET
ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024
DOI: 10.17148/IARJSET.2024.11794

containing foreground pixels (usually represented as white pixels). Consequently, the areas occupied by foreground pixels become smaller, and any internal holes within those regions tend to expand or become larger..

3.3 DILATION

The fundamental morphological operations consist of dilation and erosion. Dilation involves expanding pixels towards the boundaries of objects in an image, while erosion removes pixelsalong object boundaries. The extent to which pixels are added or removed from objects in an image depends on the size and shape of the structuring element applied during the operation.

During morphological dilation and erosion processes, the state of a particular pixel in the output image is determined byapplying a rule to the corresponding pixel and its neighboring pixels in the input image. The specific rule employed to processthese pixels defines whether the operation is a dilation or an erosion.

3.4 SILHOUETTE

Features for gait analysis are typically extracted from theGAIT ENERGY IMAGE (GEI). This process involves generating a silhouette through the following sequence of steps:

1.Conversion of video frames into individual images, followed by the transformation of color images (in RGB) into grayscale images

3.5 FEATURE EXTRACTION

A feature is defined as a characteristic or attribute of an object that can distinguish it from other objects. In the context of gait analysis, our feature vector is composed of moment features extracted from image regions that cover a walking person. Gait feature extraction is a crucial step in recognizing human gait, and it must be robust enough to handle varying conditions whiledescribing individual characteristics effectively.

3. Extraction of the moving object is achieved by performingbackground subtraction.
4. Application of image processing techniques such as erosionand dilation to enhance and refine the obtained silhouette.

The silhouette of a walking person is a promising feature to exploit since it captures the movement of most body parts and encodes both structural and transitional information. Importantly, it is independent of factors like clothing, illumination, and textures.

Given that we have a database in silhouette form, which displays most of the body parts, we can extract features from these silhouettes. There are two fundamental approaches to feature extraction: feature-based and holistic methods.

1. Feature-based methods focus on selecting individual features and characterizing geometrical relationships.
2. Holistic methods, such as principal component analysis, linear discriminant analysis, and independent component analysis, extract information from the entire image. Holistic feature extraction aims to find features with reduced dimensionality by projecting the original data onto basis vectors. These extracted features can enhance classification performance by eliminating irrelevant information from the dataset.

A feature vector is a representation of an image portion by measuring a set of features. In a 2D matrix representing an image, each individual pixel is denoted as B(i, j), which represents the brightness at the point (i, j).

During walking, the center of mass of the human body changes from one instance to another. Therefore, the center of mass is used as a feature, reflecting the weighted average of x and y coordinates of pixels within the frame. The center of mass for binary images remains the same as the center of mass if we consider intensity as a mass point. In a binary image, the centerof mass coordinates can be computed using the following formula:

IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 7, July 2024

DOI: 10.17148/IARJSET.2024.11794

$$\overline{x} = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} j * B(i, j)}{A}$$

$$\overline{y} = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} j * A(i, j)}{B}$$

IV. CONCLUSION

The integration of machine learning algorithms in breast cancer diagnosis shows substantial potential in enhancing accuracy, efficiency, and objectivity. Various studies have demonstrated the ability of ML models to analyze medical imaging data, genetic information, and clinical variables to aid in the early detection and classification of breast cancer. These advancements provide invaluable tools for clinicians and radiologists to make informed decisions, ultimately leading to improved patient outcomes.

However, challenges such as dataset quality, model interpretability, and generalization across diverse populations must be addressed to ensure the successful implementation of ML in clinical practice. Future research should focus on developing robust and interpretable ML models, exploring multimodal data integration, and considering real- time applications to further advance the field of breast cancer diagnosis. With continued research and development, ML has the potential to revolutionize breast cancer diagnosis, ultimately benefiting patients through earlier detection and personalized treatment plans.

REFERENCES

[1]. Y. Huang, D. Xu, and T. Cham, "Face and Human Gait Recognition Using Image-to Class Distance," "IEEE Transactions on Circuits and Systems for Video Technology, vol.20, no.3,pp.431-438, March 2010."

[2]. Han J, Bhanu B. B.: Individual recognition using gait energy image[J]." IEEE Transactions on Pattern Analysis & Machine Intelligence, 2006, 28(2):316-322"

[3]. Liang S C, Zhou M, An-An L I. GEI based gait recognition by using KPCA and SVM [J]. "Application Research of Computers, 2010, 27(7):2798-2800."

[4]. Theekhanont P, Miguet S, Kurutach W. Gait recognition using GEI and pattern trace transform[C]// "International Symposium on Information Technology in Medicine and Education. IEEE, 2012:936940."

[5]. Sivapalan S, Rana R K, Chen D, et al. Compressive Sensing for Gait Recognition[C]// "International Conference on Digital Image Computing: Techniques and Applications. IEEE Computer Society, 2011:567-571."

[6]. J. Han, and B. Bhanu, "Individual recognition using gait energy Image, IEEE Transactions on Pattern Analysis and Machine."