ISSN 2228-9860 eISSN 1906-9642 CODEN: ITJEA8

International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies

http://TuEngr.com

Towards Autonomous Micropipette Positioning in Eye Surgery by Employing Deep Learning Algorithm in Micro-Cannulation

Mukesh Madanan[1*], Nurul Akhmal Mohd Zulkefli[1]

[1]Department of Computer Science, Dhofar University, Salalah, OMAN.
**Corresponding Author (Tel: +968 93681309, Email: mukesh@ du.edu.om).*

Paper ID: 14A1A

Volume 14 Issue 1

Received 23 July 2022
Received in revised form 19 October 2022
Accepted 26 October 2022
Available online 01 November 2022

Keywords:
Artificial Intelligence;Machine Learning;Deep Learning;Robotic Surgery;Eye Surgery;Micro-cannulation;Enhanced Guassian Filtering;Bee Colony Optimization;CNN;Image Processing

Abstract

Eye surgery, more precisely the retinal micro-surgery involves both sensory as well as motor skills. This is confined within human boundaries along with physiological limits for maintaining consistent steadiness, the ability to feel small forces and accuracy. Despite these assumptions to leverage robots in all types of surgery, multitudes of challenges have to be confronted to reach complete development. The deployment of robotic assistance in ophthalmologic surgery also faces the same challenge. This work focuses on the autonomous positioning of a micropipette that is to be mounted on a surgical robot for performing eye surgery. Initially, multiple microscopic images of the given micropipette along with its shadow are collected. These images are treated or filtered by using the Enhanced Gaussian Filtering (EGF) method. The so-obtained filtered image is partitioned or segmented by Bee Colony Optimization (BCO) into three segments: micropipette, eye ground and shadow of the micropipette. A new Modified Convolutional Neural Network (MCNN) is leveraged by the robot to perform eye surgery that learns the microscopic images with their ground truth. This MCNN uses automatic feature extraction and estimates micropipette regions with their shadow by examining a microscopic image and its tip. This is tapped for developing autonomous position control in robots. The selected micropipette is found to be positioned at a 99.56% success rate with a mean distance of 1.37 mm from the eye ground that is simulated.

Discipline: Artificial Intelligence, Healthcare, Applied Information Technology

Cite This Article:

Madanan, M., Zulkefli, N. A. (2023). Towards Autonomous Micropipette Positioning in Eye Surgery by Employing Deep Learning Algorithm in Micro-Cannulation. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies, 14*(1), 14A1A, 1-20. http://TUENGR.COM/V14/14A1A.pdf DOI: 10.14456/ITJEMAST.2023.1

1 Introduction

1.1 Information

Eye surgery, more precisely retinal micro-surgery, necessitates sensory as well as motor skills that mimic human boundaries along with physiological limits with consistency, accuracy, steadiness, and the potential to isolate the minute forces that are involved in the operation [1]. Besides the benefits of deployment of robots in medical surgery and despite the efforts placed for further development, this field faces a multitude of challenges to impart robots in real-time. The field of surgical ophthalmology also suffers from the same issue. The first robot-assisted ophthalmology-based retinal surgery was performed after 30 years after the publication of the first experimental papers pertaining to the subject that too in-human presence [2]. Artificial Intelligence (AI) through sourced a few decades ago, but it has now emerged as indispensable technology, especially in all surgeries and ophthalmology. This time gap was mainly due to the hampering of the technological advancements to realize AI as a tool for implementing novel processing methods.

The primary technique among technological advancements is the processing prowess of Graphics Processing Units (GPU) for Machine Learning (ML) [3]. It is evident that the conventional notion of robots is to perform repetitive tasks mostly, but AI and ML are closely related with proven abilities to design new concepts to resolve problems autonomously through perceiving and understanding. The ultimate implication of exploring these abilities is that future machines may intricate into the domain which was once delineated as human-reserved [4]. Though the capacity of AI/ ML can be witnessed strongly, the current marketing promises as well as the hype created outnumber its pace of development. However, these robotic systems are integrated with ML that significantly enhancing the quality and precision of robot-assisted retinal surgery, which can be seen as a major breakthrough in the discipline. In addition, a lot of research comprehensively analyzes the recent technological advances in the domain of retinal robotic surgery by throwing light on the bottlenecks and limitations, apart from briefing the rationale of AI in robot-assisted retinal surgery [5].

As a matter of fact, particularly in retinal microsurgery, it has been discovered that human precision is best between 20 and 40 m. According to reports, the average human tremor has a peak-to-peak value of about 100 m. The force indicated in the earlier work that is sufficient to tear the rabbit's retina is close to 7.5 mN [6], which is the mean human threshold for increased tactile sensibility. In this case, the stability, accuracy, and sturdiness of the robot are quite helpful for handling delicate intraocular surgeries. The development of neural networks [7], their improvements in image processing, and the notable rise in data utilization [8] may eventually increase the efficiency and safety of robotic techniques used in eye surgery.

Apart from rendering new data sources, surgical visualization systems like “heads-up surgery” integrate Augmented Reality (AR) during the surgery. Virtual Reality (VR) systems like

Eyes Surgical as well as Deep Learning (DL) tracking methods can also be leveraged with robotic control. This will definitely aid in training, improving and testing these systems to eliminate iatrogenic injuries. The improved data is the pedestal of progress in advancing AI, improved safety, enhanced efficacy and inclined reliability which are some of the potential outcomes of the integration of robotics into the discipline of ophthalmology. Several barriers exist to end-user adoption of robotics not confined to size, costs, functional limits, human acceptance, accuracy and more importantly, yielding better outcomes with safety.

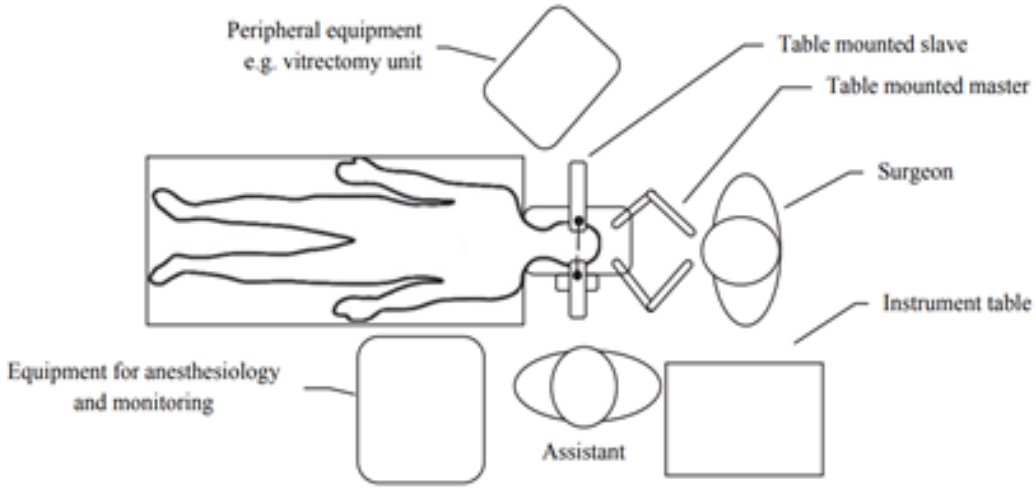

Figure 1: A typical operation room layout for eye surgery-Master Slave surgical process

With this motivation, this work proposes deep learning-based position finding of the micropipette attached to the surgical robot during eye surgery. Eye surgery, or in particular, vitreoretinal eye surgery, is an application where a manually or tele-operated robotic system assists the surgeon in performing the surgical routine. This robotic system is desired to be a master-slave system, where the slave system performs the actual surgery, controlled by the surgeon at the master as shown in Fig.1. As such, it can bring stability to enhance the surgeon's surgical skills, while the surgeon's knowledge and experience can still guide the process. This work presents some background information on robotics and describes some specific applications in medicine and vitreoretinal eye surgery.

It is preferred to maintain the current operating room layout, where the patient and peripheral equipment is within arm's reach. A table-mounted slave system is compact and contributes to this [9].

The paper deals with an autonomous vitreoretinal eye surgery robot along with the accuracy-enhancing approaches to reach the target on eye ground. This robot ceases to be switched to the autonomous injection, which is also to be developed as a part of the work. The primary contribution of the work is as follows:

- Describing the working of autonomous positioning of a given micropipette along the depth direction by tapping its shadows where the image pre-processing is done using enhanced Gaussian (EGF).
- Segmentation of the given micropipette position and shadow of the micropipette, along

with eye ground regions is done using Bee Colony Optimization (BCO).

- Finally, automatic feature extraction and the prediction of a label based on a Modified Convolutional Neural network (MCNN) for determining the shadow region that has an unclear border in the image, where stochastic max-pooling was injected to replace the standard average stochastic maximum pooling and maximum stochastic maximum pooling used.
- The positioning of the micropipette for eye surgery assists the robot automatically, which is done through a visual servo control method augmented with planar motion.
- The suggested method is particularly practical since, during vitreoretinal surgery, in particular, surgeons frequently use shadows to place tools along the depth axis. Astigmatism has little effect on how shadows are used in the end.

This paper is organized as follows. Section 2 describes the related work on eye surgery robots using ML and DL methods. Section 3 proposes Autonomous Positioning of the micropipette tip using the MCNN technique. Section 4 details the experiment that is done to assess the proposed method with discussions on the results, and the conclusion and future research directions are briefed in Section 5.

2 Literature Review

This section elaborates on the usage of ML and DL as a part of modern surgery, with a special focus on AI issues. The advancements of ML and DL will enable surgeons to deploy more autonomous actions in the surgical process. Tayama et al., [10] developed a novel method of automatic positioning of robotic arms by tapping the shadow of surgical instruments involved in the surgery. The microscopic image is partitioned into 3 regions: a micropipette, shadow of the micropipette, and eye ground through the famous Gaussian Mixture Model (GMM) augmented with Kalman filters. Li et al., [11] designed low-cost yet robust Densely connected CNN (Dense CNN) that could sense 9 or 36 directions of gaze computation to control the surgical robots. The surgeons can effectively control the developed robot by mentioning the beginning as well as ending points of the surgical robot through an eye gazing mechanism. These surgical robots can be managed to progress in 9 directions by obtaining the information from the controllers' eyes gaze.

Li et al., [12] proposed a probe-based confocal laser endomicroscopic system (pCLE) which acts as an imaging modality for enhancing the diagnosis. The capability to view the retina at a more granular cellular level provides information that is important to foresee the surgical results. Enhanced image quality, and smoother motion, with the decreased workload, can be achieved with hybrid control frameworks. But few researchers concentrated on it during the surgery. This is indeed a big challenge due to the high deformation of the iris and the occlusion that is inherently caused by instruments that are used in the surgery. A new real-time iris tracking system that is based on a regression network is given by Qiu et al., [13] that meets the accuracy as well as speed requirements demands of the ophthalmic robotic system. This system employs low-

level visual features with high-level semantic meanings that are collected from different layers to extract and capture the discriminative representation of the target (iris).

The method for tracking objects involves looking through the windows with the best classifier scores. To locate the iris center, the Scale Invariant Feature Transform (SIFT) [14] is used [15]. Through PCA, a model in the shape of an eyelid is created and used to follow a particular iris [16]. According to some experts, template matching is still an effective method for identifying the iris. The feature extraction from photos is constrained by the feature representation's incapacity to handle more complicated real-world settings. The surgical scene typically contains more occlusion of the actual surgical tools being utilized. Consequently, the handcrafted features won't be enough to monitor the iris for the creation of an ophthalmic robotic system.

The features extracted from CNN make a bigger contribution to state-of-the-art object trackers. These features display a very strong representation of inherent semantic information of the pertaining target. This has indeed attracted greater research interest. Li et al., [17] designed a cascaded CNN to isolate the eye from facial images. The CNN can categorize the underlying region as left/right eye and the next CNN is employed for detection. Harini used ensemble learning in ResNet10 to track the eye [18]. Wolfgang *et al.* [19] suggested a double CNN for the detection of pupil position. These proceeds in two phases, the pupil position are identified by the first CNN while the second is used to refine its position.

Hoffman et al., [20] proposed a CNN for eye detection, which incorporates a segmentation to learn the relative significance of the pupil as well as the iris. Though the DL methods outperform the hand-crafter and correlation filter methods, most network-based trackers are incapable of achieving real-time tracking mainly because of the online training. The multi-loss objective function is constructed to obtain improved overlap precision. These modifications eventually enhance the accuracy of tracking in the underlying ophthalmic robotic system.

In order to predict the occurrence of negative occurrences from a relatively brief history of temporal sensor readings, He et al. [21] introduced a Recurrent Neural Network (RNN). The robot is managed with a variable admittance controller to prevent it from tapping into undesirable situations. Through microscope-integrated Optical Coherence Tomography (MI-OCT), Zhou et al. [22] demonstrated a more reliable framework designed for needle identification as well as localisation relevant to the sub-retinal injection. The results show that the suggested strategy can localize the needle more precisely with a confidence level of about 99.2%. The already existing ML posts multiple challenges, like the demand to process a voluminous amount of data, iterative model training, and tuning the model according to various scenarios. But with the advent of DL, it is definitely possible to effectively train an end-to-end DL model to resolve complex tasks.

3 Method

The work presents a more generic and robust framework for detecting the needle and localization of the same in the process of sub-retinal injection by examining microscopic images using DL. This method progresses through three main steps: image pre-processing, segmentation

and locating the needle in segmented images. Pre-processing of the microscopic images is done through EGF. Then the segmented image fragments are fed into BCO for detection of the needle with its localization. The developed MCNN estimates the regions of the micropipette and its tip positions along with its shadow from a real-time image. The planar positioning is done through the visual serving method [24]. The utilized micropipette is advanced to the eye ground till the distance of the micropipette's tip and its corresponding shadow is less than or equal to the already predefined threshold. Therefore, the robot could effectively and accurately reach the eye ground and also stop safely before it establishes contact with the surface. The experimental results indicate that this method can vividly localize the needle with a good confidence score of 99.2%. The generic framework of the fore mentioned methodology is given in Fig.2.

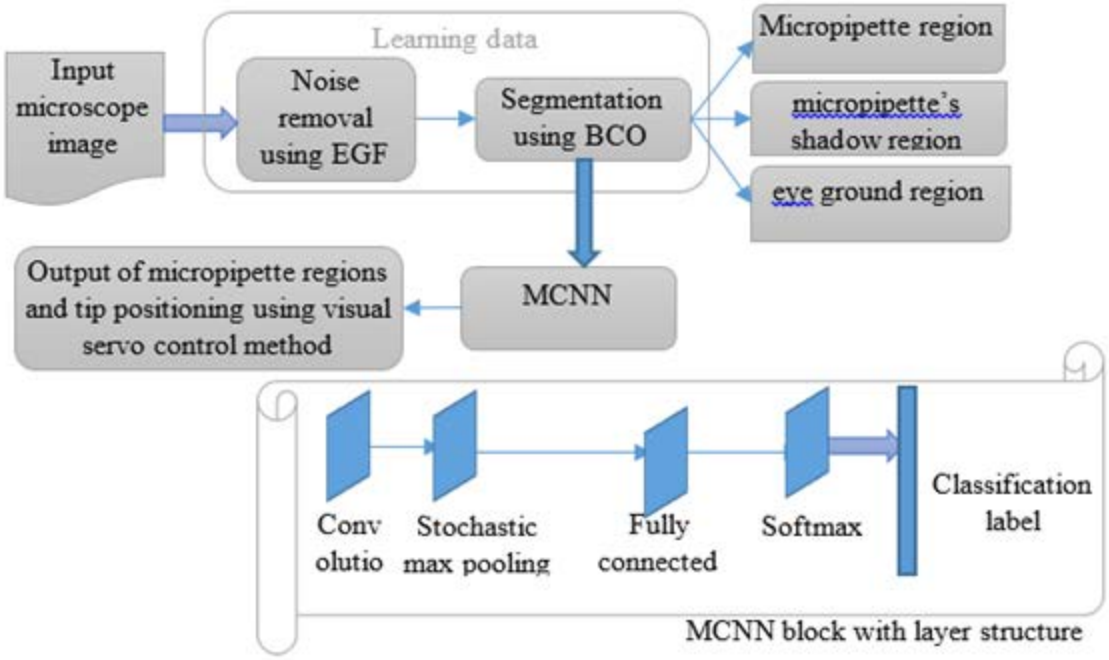

Figure 2: General framework diagram of the proposed methodology

3.1 Image Enhancement Using Enhanced Gaussian Filtering

Multiple microscopic images of the micropipette along with its shadow are obtained by varying the position of the micropipette and ambient conditions. After this, these images are segmented and the components are labelled. For better visualization, the micropipette, its shadow, and its background are rendered in different colors namely blue, red, and green, respectively. The pairs of microscopic image with ground truth data are employed in preoperative learning through the MCNN model. The EGF isolates the pixels with gray levels ranging from 0 -255 as noise. It restores the noise pixels through a Gaussian filter by estimating the noise density. Table 1 enumerates the parameters used in the proposed method.

Table 1. Parameter and Description of Proposed Methodology

Parameters	Description
$P_{i,j}$	microscopic images pixel, $i, j = 1, \dots, L-1$
$g_{i,j}$	Gray level of $P_{i,j}$
Ω	$\omega \times \omega$ filtering window
$Res_{i,j}$	the restored gray level of $P_{i,j}$
$P_{m,n}$	neighboring pixel of $P_{i,j}$
$\varpi_{m,n}$	weight of $P_{m,n}$
σ^2	adaptive variation
nd	noise density
$s_{ij} = x_{ij} + \Re_{ij}(x_{ij} - x_{kj})$	BCO swarm search
k	Food source $k = \{1,2, \dots, NS\}$
NS	number of food sources
x_{kj}	randomly generated k food source
$\Re_{ij}$	random number
$(\dot{m}, \dot{n})$	optimum 2-dimensional vector of the objective function
$\varphi(m,n)$	two-dimensional entropy
E_{m}	two-dimensional entropy of micropipette
E_{e}	two-dimensional entropy of eye ground regions
(m,t)	threshold value
R	Region
RP_1 and RP_2	probabilities of Regions 1 and 2 respectively
MC	maximum cycle number
D	Population size on search space
X_i	New solution
x_i	feature vector
l_i	Label of microscope image
$layer$	serial number of the layer
RF_j	the receptive field of the input layer
y_{max}	Result of pooled maximum value
y_{avg}	result of average pooling
Prb_i	Probability for value v_i
$y_{stochastic}$	result of stochastic pooling
GL_c	global loss of classification
GL_r	global loss of regression
Prb_i^*	the indicative value that corresponds to i-th candidate detection box
ap_i	coordinate vectors of the micropipette's detection box
ap_i^*	coordinate vector of the tip point that corresponds to the calibration box that depicts the shadow
λ	Threshold value
N_r and N_c	Normalized regression and classification respectively

3.1.1 EGF

For any given microscopic image with a pixel $P_{i,j}$ with gray level $g_{i,j}$, the EGF method isolates the noisy pixels for the specified i *and* $j = 1, \dots, L-1$. If the $g_{i,j}$ is 0 or 255, then the EGF decides the pixel $P_{i,j}$as noise. Otherwise, the $P_{i,j}$ is delineated as noise-free. This noise isolation method allows reduces the wrong detection to 0. But sometimes it may detect noise-free pixels pertaining to the gray levels 0 as well as 255 as noisy, thereby generating false positives. This miss detection permits the noise-free pixels to undergo unfavorable filtering that may result in a blurred image. Apart from this, the inaccuracy of this method that is caused by the false alarm may adversely

affect the performance of noise density estimation as well as the process of adaptive variance determination embedded in the Gaussian filter that is commonly used as a noise restoration technique.

The fore mentioned EGF method replaces the noisy pixel with its weighted average of the filtering window of fixed size. These weights are determined by a Gaussian filter augmented with the adaptive variance. The comprehensive noise restoration process in the EGF is given as:

For any detected noisy $P_{i,j}$, Ω indicates $\omega \times \omega$ filtering window that is centred at $P_{i,j.}$. The value of ω is set to 9.

Form a fresh set $\mathcal{N}$ excluding the noisy pixels at gray levels 0 as well as 255 at Ω. Compute the weighted mean of the pixels as restored gray level ($Res_{i,j)}$.

$$Res_{m,n} = \frac{\sum_{P_{m,n} \in \mathcal{N}} \varpi_{m,n} * g_{m,n}}{\sum_{P_{m,n} \in \mathcal{N}} \varpi_{m,n}} \quad (1),$$

The terms $g_{m,n}$ refers to the gray level and $\varpi_{m,n}$ is the weight of $P_{m,n}$. The coordinates (i,j) as well as (m,n) respectively indicates $P_{i,j}$ and it's neighbor $P_{m,n}$.

Weight ($\varpi_{m,n}$), is computed using the Gaussian function as:

$$\varpi_{m,n} = \exp\left(-\frac{(m-i)^2+(n-i)^2}{2\sigma^2}\right) \quad (2),$$

The parameter σ is also empirically set as

$$\sigma = nd + 0.2 \quad (3)$$

σ^2 is the adaptive variance of a given Gaussian function. The term nd indicates the noise density which is computed as the ratio of the count of detected noisy pixels with a cumulative count of pixels in the entire image.

3.2 Segmentation Using BCO

The images that were pre-processed are categorized or segmented into micropipette, the shadow of the micropipette, and eye ground regions. Every pixel of the regions is labelled to be deployed as a ground truth. The proposed MCNN is employed by an eye surgery robot to study the microscope images with their ground truth through HSV color as features. The performance of this segmentation completely depends on local as well as global searching capability. Hence, the method of collecting honey forms the motivation for BCO and is given as $s_{ij} = x_{ij} + \Re_{ij}(x_{ij} - x_{kj})$. The term x_{kj} is randomly generated where $k \neq i$ (food source). The value of $k = \{1,2,\dots,NS\}$. NS indicates the total count of food sources available. $\Re_{ij}$ is assigned a random value within $[-1,1]$. The search step of the insect is estimated as $(x_{ij} - x_{kj})$. The food source (k) is chosen around its surroundings. The popular Levy flights are injected to aid the foraging behaviors of bees to improvise the development of the algorithm and also to reinforce its capability to move away from local extreme values. Also, BCO uses global optimal solutions as well as individual extremes as a part of searching in observation bees. This collaboration among the populations adds up to global convergence. When the values$(\dot{m},\dot{n})$, are maximum, then the gray level pixels of the object as well

as its background will be uniform. The objective function for finding the optimal threshold $(\dot{m},\dot{n})$ is taken as the popular maximum 2D entropy threshold as given by:

$$(\dot{m},\dot{n}) = \arg\max_{0\le m\le L-1} \max_{0\le t\le L-1}\big(\varphi(m,t)\big) \quad (3),$$

This algorithm tries to achieve this value that can effectively maximize $\varphi(m,t) = E_{\text{m}} + E_{\text{e}}$. The value E_{m} is the 2D entropy of the micropipette. The value E_{e} is 2D entropy of eye ground regions.

$$E_{\text{m}} = -\sum_{i=0}^{m}\sum_{j=0}^{t}\frac{RP_{ij}}{RP_1}\ln\left(\frac{RP_{ij}}{RP_1}\right) \quad (4),$$

$$E_{\text{e}} = \sum_{i=m+1}^{L-1}\sum_{j=t+1}^{L-1}\frac{RP_{ij}}{RP_2}\ln\left(\frac{RP_{ij}}{RP_2}\right) \quad (5),$$

A 2D histogram can be constructed using 2D entropy modelling. The fixed threshold value (m,t) is partitions the 2D histogram into 3 rectangular zones namely: Regions 1-3. Region 1 pertains to the micropipette while Region 2 represents its shadow. Region 3 clearly indicates the ground eye regions. Generally, region 2 is ignored. The estimated probabilities of the first and second regions are given as $RP_1 = \sum_{i=0}^{m} p_{ij}$ and $RP_2 = \sum_{i=m+1}^{L-1}\sum_{j=n+1}^{L-1} p_{ij}$, respectively. A small difference in the gray level pixel yields great 2D entropy. If the 2D entropy is large, then the gray-level distribution within a class is assumed to be more uniform. This work attempts to maximize the 2D entropy of the target along with its background classes to achieve segmentation. The complete flow diagram is shown in Fig. 3. The BCO intends to resolve maximization problems. Hence, its objective function is maximized as the reciprocal of $(\dot{m},\dot{n})$. The complete pseudocode of the proposed BCO is presented below.

Input: Pre-processed images as the number of source of bee NS, the size of the population is given as D. The solution's dimension is D. The count of the honey-collecting and observation bees are set as initial iterations with Cycle = 1 whose maximum value is MC

Begin

Initialize the parameters pertaining to the population. Eliminate solution as well as upper bound of iterations MC.

Assessing x_{ij}, $(i = 1,2,\dots,NS, j = 1,2,\dots,D)$ at Cycle = 0, limit = 0.

Do

// honey collecting bees

For the value i =1 to NS then do

generate the new solution

$X_i = X_{min} + rand(0,1)(X_{max} - X_{min})$ and $fitness_i = f(X_i)$

Evaluate $x_{ij}(n)$.

If $(x_{ij}(n+1)$was superior to $x_{ij}(n))$

replace by $x_{ij}(n)$ by $x_{ij}(n+1)$

Else

Increase the iteration number by 1

End If

End For

Record the optimal or best solution X_{best}.

//observation bees

For the value $i = 1$ to NS

If the condition (rand < RP_i) holds

Generate the new solution

$X_{ij}(n+1) = X_{ij}(n) + \text{ꞩc} \otimes Levy(\lambda)$,

// $\text{ꞩc} = 0.01$ is the amount of the step control and $1 < \lambda \leq 3$

Evaluate $x_{ij}(n)$.

If ($x_{ij}(n+1)$ was superior to $x_{ij}(n)$)

replace $x_{ij}(n)$ by $x_{ij}(n+1)$.

else

Increase the iteration number by 1

End If

End For

// scout bees

For the value $i = 1$ to NS

If (iteration ⩾ MC)

Discard solution, sp_i. This can be employed to get a new solution.

$sp_i = \frac{fitness_i}{\sum_{n=1}^{NS/2} fitness_n}$//where $fitness_i$ is the fitness of food sources.

End If

End For

Record the best solution X_{best} till now, MC = MC + 1.

Through the obtained wo-dimensional entropy threshold for segmentation

Until MC is reached

End.

Assigning the present but optimal value as well as depending on the relationship among the populations, the algorithm improves convergence speed. Adoption of a new fitness assessment and integration with 2D entropy-based multi-threshold segmentation, the proposed modified BCO tests the threshold segmentation.

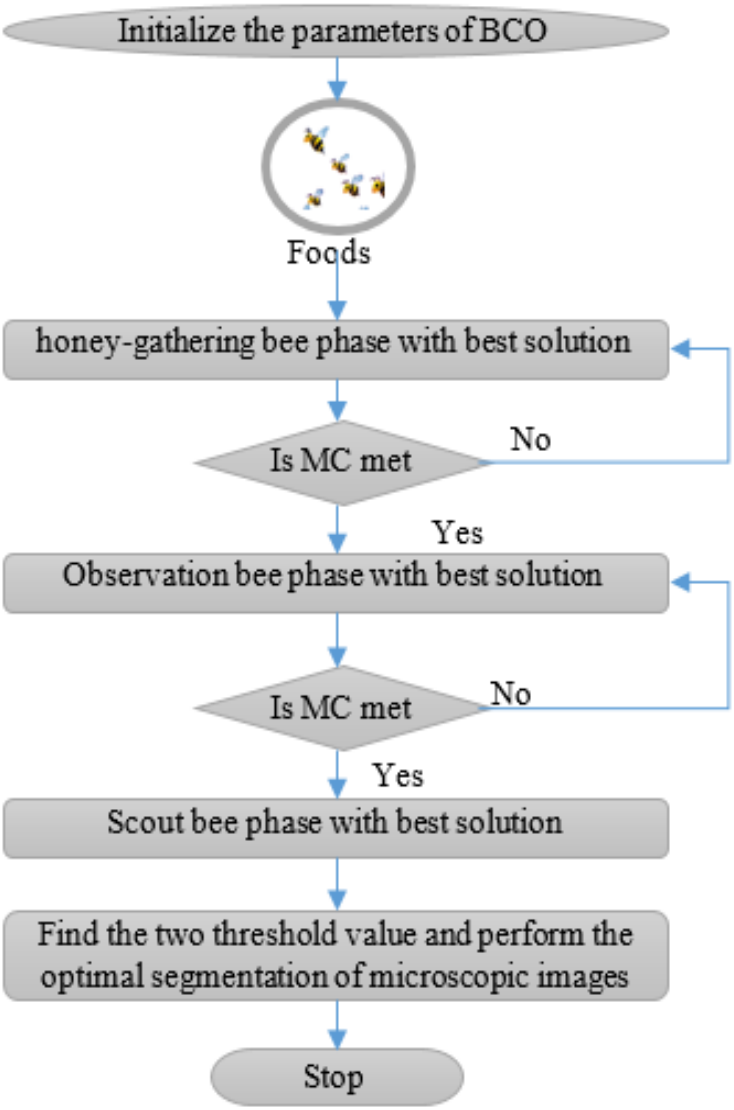

Figure 3: Flow diagram of BCO for segmentation

The feature vector (x_i) of i-th pixel of every image is mentioned as l_i. This is in accordance with the corresponding pixel of the ground truth image, mentioned as [10].

$$x_i = (x_1, \dots, x_k) \quad (6),$$

$$l_i = (l_1, \dots, l_k) \quad (7),$$

The term x_i represents Hue, Saturation, and Value (HSV) of i-th pixel of the image. The value $k = 954486$ indicates the count of pixels in the image. The MCNN model forms the feature vectors (x) along with its labels l . The computed label of the pixel in the image is given rooted in MCNN learning.

3.3 Modified Convolutional Neural Network for Autonomous Robotic Position Control

The proposed work partitions the input data as testing and training sets. During the training phase, the model consumes the already processed images. Whereas in the testing phase, the model evaluates the classifiers by validating them on the testing set. The model labels the image as

normal or an attack. In order to develop an MCNN-based intrusion model takes each labeled data into 13×6 size images because each data contains 29 features except the 'Label' feature which is used for image classification. MCNN is primarily construed with RPN along with stochastic max pooling. A more simplified VGG16 is used for feature extraction. The RPN finds if candidate box if the image comprises the target. The primary aim of MCNN is object detection. The structure of MCNN is portrayed in Fig.4.

To extract the feature figure a fundamental feature extraction network is employed that delves HSV features from the image. The VGG16's fully connected layer is eliminated, and stochastic max pooling is chosen to extract the basic features. This version of VGG16 based feature extraction network has convolution along with a stochastic max pooling layer augmented with it. These features cannot be delineated as just primitive features like color and shape. As a matter of fact, the feature law remains unclear. Therefore, isolating the tip along with its types demands much more advanced abstract features. The convolution operations must ensure that overall image features are not overlooked. Also, the network depth must meet the requirements of finding abstract features. Hence, the complete network increases the depth of the underlying network which eventually improves the learning as stacking multiple 3×3 convolution kernels will definitely enhance the efficacy. To combat over-fitting as well as to improve generalization, stochastic max pooling is applied to all pooling layers. The dimensions of the image namely the length and width, are decreased to half after every pooling layer. This results in the formation of smaller yet concentrated and abstract feature images.

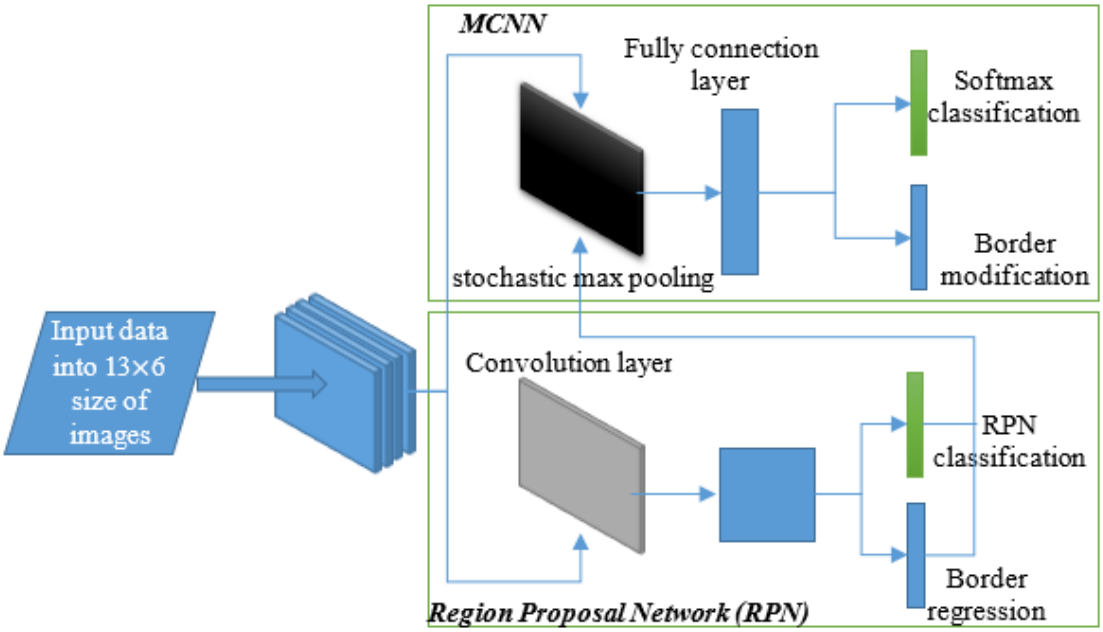

Figure 4: The overall structure of MCNN.

3.3.1 Convolution Layer

In this layer, each neuron is connected to the local receptive field of the previous layer. The local features are obtained by the convolution process. The complete convolution can be given as:

$$x_j^{layer} = f\left(\sum_{i \in RF_j} x_i^{layer-1} \times ker_{ij}^{layer} + bias_j^{layer}\right) \quad (8).$$

The *layer* indicates the serial number and *ker* represents the convolutional kernel. The receptive field is represented as RF_j.

3.3.2 Stochastic Max Pooling

This ranks all receptive field values and randomly selects top-n values. This method is different from the conventional average, max and stochastic pooling strategies. This is actually a combination of max and stochastic pooling. As an initial step define 3 generic pooling strategies. Max pooling, yields the maximum value pertaining to each receptive field and is given as

$$y_{max} = \max(RF) \quad (9),$$

The term y_{max} is pooled maximum value while RF represents the values of the receptive field. The average pooling operates by computing the mean of all values in the receptive field and is given as:

$$y_{avg} = \text{mean}(RF) \quad (10),$$

The value y_{avg} is the outcome of average pooling. Generally, the stochastic pooling chooses a value from the receptive field with some probability (Prb_i) at value (v_i)at position (i) and is given as:

$$Prb_i = \frac{v_i}{\sum_{k \in RF} v_k} \quad (11),$$

The common stochastic pooling is given as:

$$y_{stochastic} = v_l, \text{ where } l \sim Prb(prb_1, \ldots, p_{rbl}) \quad (12),$$

The term $y_{stochastic}$ gives stochastic pooling. Here, the pooling result is chosen from top-n values.

3.3.3 Region Proposal Network (RPN) Model

The input image is transformed into a feature map by subjecting it to the fore-mentioned basic feature extraction network. This feature map is scanned by using a 3 × 3 sliding window. The eigenvector of size 512 will be generated after every scan. This is then sent to fully connected layers. After this, 2k as well as 4k convolution kernels each of size 1 × 1 are employed to perform feature mapping of both classifications as well as regression layers with a step size of 1. This anchor point forms the pivotal sliding window point. Every sliding window forms 9 prediction windows with 3 kinds of size ratios and length: width ratios of 1:1, 1:2 as well as 2:1. The value k indicates the maximum number of prediction windows. Hence k is fixed as 9. These windows must be discarded and then selected. The prediction window that falls beyond the boundary of the image will be discarded. The original label box of input data along with the remaining prediction boxes is

used to estimate the ratio of overlapping elements. The prediction box that has an overlap ratio within the range [0.3, 0.7] will be eventually discarded.

3.3.4 MCNN and RPN

The network under study is trained by employing an alternating training method. Since the model has scanty data, it will not be advisable to start the training from the scratch. Therefore, initialization is done using ImageNet parameters. The RPN is trained individually to form a suggestion box. The parameters of this network were adjusted from end to end. The training data is trained by MCNN detection, from which the suggestion boxes are eventually generated by fore mentioned RPN. The MFCNN initializes the RPN based on fixed but shared convolutional layers. The RPN outputs the candidate box that was used earlier used as an input to modify the parameters of MFCNN. The global loss (GL) is estimated as:

$$GL(Prb_i, ap_i) = \frac{1}{N_c}\sum_i GL_c(Prb_i, Prb_i^*) + \lambda\frac{1}{N_r}\sum_i Prb_i^* GL_r(ap_i, ap_i^*) \quad (13).$$

$$GL_c(Prb_i, Prb_i^*) = -\log[Prb_i^* Prb_i + (1 - Prb_i^*)(1 - Prb_i)] \quad (14).$$

$$GL_r(ap_i, ap_i^*) = smooth\ L1(ap_i, ap_i^*) \quad (15).$$

The term i holds an integer value. Prb_i is the likelihood that i-th candidate detection box is rightly predicted to be the target. Also, the Prb_i^* gives an indication value that correlates to i-th candidate detection box. If the same candidate detection is the underlying target object then the value of $Prob_i^*$ is maintained as 1. If the same is a background, then Prb_i^* assumes the value of 0. The term $ap_i = \{apx, apy, apw, aph\}$ gives 4 parameterized coordinate vectors pertaining to the detection box. Whereas ap_i^* $\{apx, apy, apw, aph\}$ represents the coordinate vector of the tipping point that corresponds to the calibration box. The value $GL_c(Prob_i, Prob_i^*)$ refers to the loss during classification while $GL_r(ap_i, ap_i^*)$ is the regression loss. The value N_c expresses the value when the normalized value of the underlying classified item adheres to the min batch. The value N_r is the regression term that is eventually normalized to the count of tip positions, especially in eye surgery robots. The above-mentioned micropipette is advanced autonomously with a finite but constant velocity. This system moves the robot by estimating the distance between the micropipette tip and its corresponding shadow. The robot stops its movement when the estimated distance is less than or equal to a predefined threshold. This threshold can sometimes be roughly zero. This often occurs when the estimation of tip positions of the micropipette is accurate.

3.3.5 The Automatic Tip Position Estimation

The real-time image is processed, to find the tip as well as its shadow by subjecting it to MCNN. The present tip position is transferred to the control PC. This micropipette was managed and controlled to be accurately placed at an already well-defined target in microscopic view through the famous visual serving [23]. Consequently, the robot advances toward the eye ground till the gap between the tips of the micropipette and its corresponding shadow declines below a

predefined threshold level. The term ap_{target} in the already defined microscopic image with velocity given as

$$\dot{ap}_{target} = \gamma J^{+}(ap_{target} - ap_t) \quad (16),$$

The term γ is control gain and ap_{target} is the image feature of the target position obtained from the micropipette. The ap_t is an image feature of the pertaining robotic position. t refers to the timestamp. The so observed micropipette progressed automatically with fixed velocity. The system under study moves the robot. It also calculates the magnitude of the distance between the micropipette tip and its corresponding shadow. The robot stops moving when the distance is less than or equal to an already defined threshold. This threshold can even be roughly zero when high accuracy is reached.

4 Results and Discussion

The experiment to assess the learning data is done with 5-fold cross-validation [10]. Five pairs of images with ground-truth data are used by altering the positions as well as the strength of the light. As a part of the experiment, halogen light is employed for illuminating the micropipette. A customized force sensor is positioned under the eye ground to find the contact of the micropipette's tip. Next, 4 pairs of microscopic images along with the ground truth of Experiment I are employed as learning data for the so-developed MCNN. The target in the image was fixed by selecting the corresponding pixel. The proposed methodology MCNN is compared with the existing method of GMM [10] and Dense CNN [11] with the performance metrics of detection accuracy, recall, f-measure and precision.

4.1 Evaluation Metrics:

The equations of the evaluation metrics are given as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \times 100 \quad (17).$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \times 100 \quad (18)$$

$$\text{F} - \text{measure} = 2 * \left(\frac{\text{recall} * \text{precision}}{\text{recall+precision}}\right) \quad (19)$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \times 100 \quad (20)$$

The term TP indicates True Positive while FP refers to False Positives. The True Negative is indicated as TN whereas False Negative is denoted by FN.

4.2 Precision Rate Comparison

The accuracy of suggested and current models for the number of features in a given database is shown in Fig. 5. While the number of characteristics grows, the corresponding precision also increases. For instance, the MFCNN offers 94% more precision than the GMM and Dense CNN. This is so that the MFCNN can find a relatively better-sorted collection of input within a given period of

time without requiring high-dimensional features or derived factors. The proposed method was able to attain improved results since the dataset possesses images with high dimensions. The system can effectively handle as well as enhance the positioning of the tip during eye surgery.

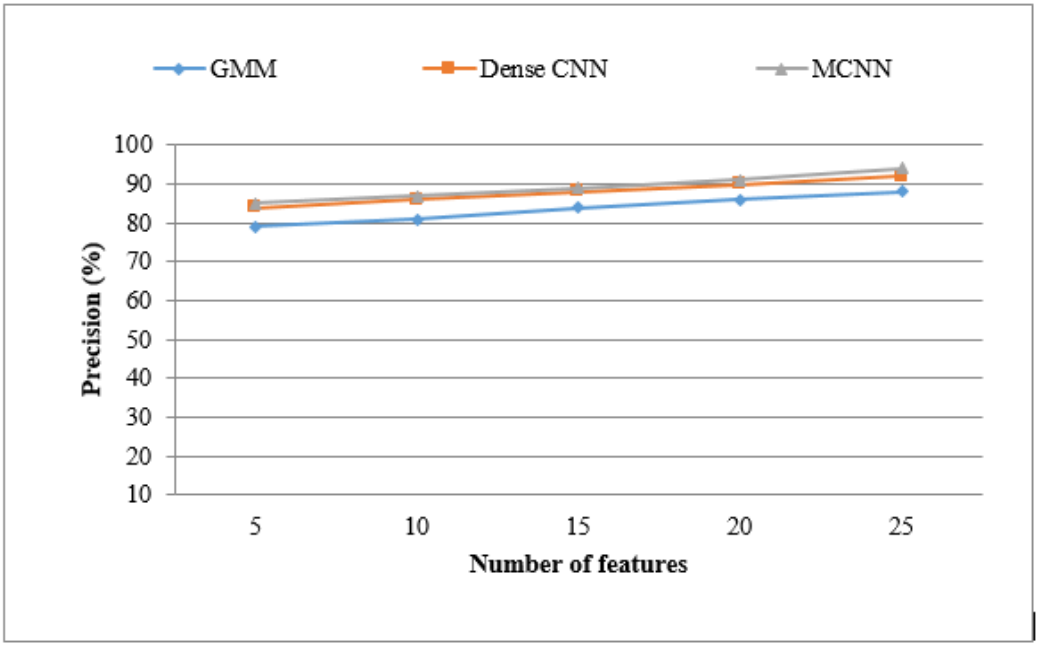

Figure 5: Result of Precision

4.3 Recall Rate Comparison

The recall of proposed and existing models for the number of characteristics in a particular database is shown in Fig. 6. Recall is maximized as the number of features increases. Comparing the MFCNN to the GMM and Dense CNN, for instance, a recall of 92% is achieved. This is due to the fact that the BCO reduces the calculation time of the resulting factors, allowing for the precise segmentation of three regions for subsequent MCNN-based tip position determination. The proposed DL algorithms show supreme results in detecting the tip positions of the underlying eye surgery robot.

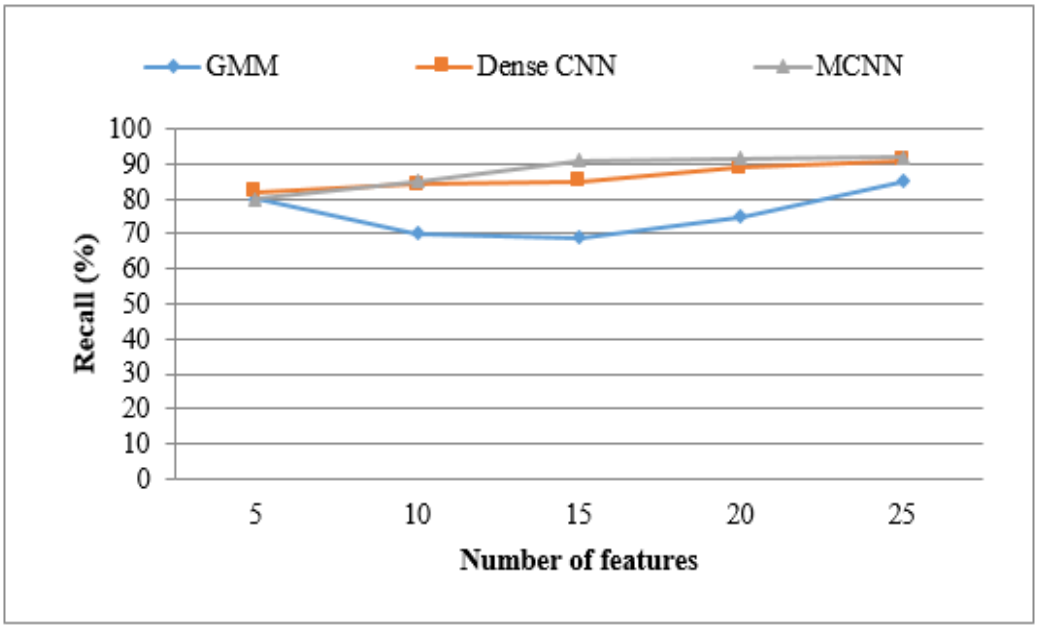

Figure 6: Result of Recall

4.4 F-Measure Rate Comparison

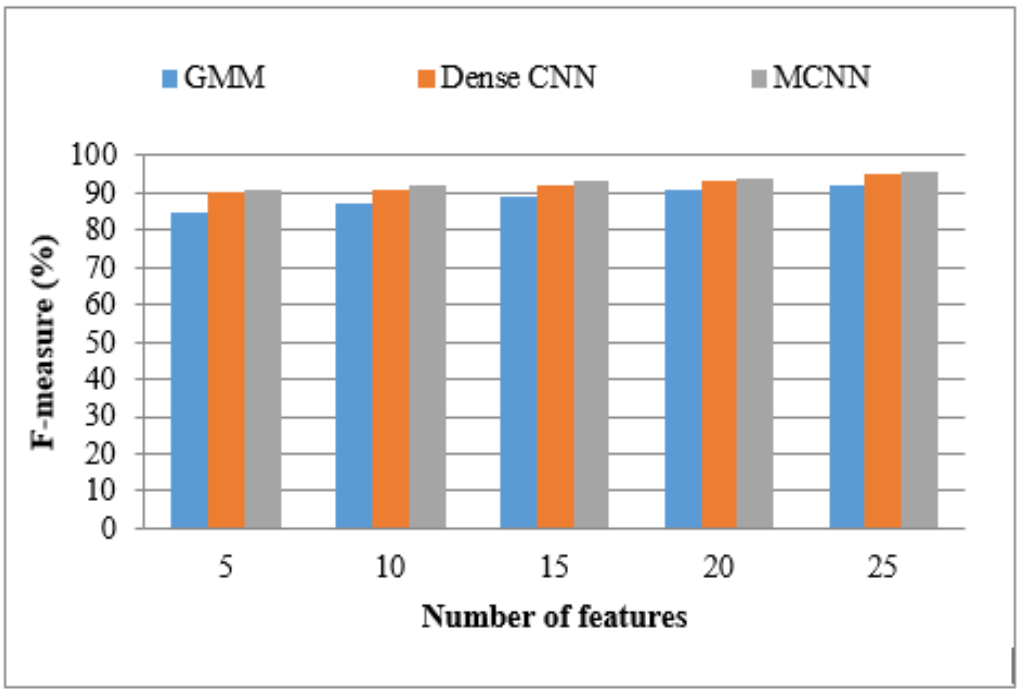

Figure 7: Result of F-measure

The f-measure of proposed and current models for the number of features in a specific database is shown in Fig. 7. The f-measure is maximized while the number of features is maximized. The MFCNN, for instance, has an f-measure of 96% when compared to all other models. EGF algorithm has been applied for eliminating the noise, thus improving image quality. Therefore, the further processing, memory requirement, and time complexity can be reduced by MCNN with the advantage of automatic HSV feature extraction, especially for automatic and autonomous positioning of the robot.

4.5 Accuracy Comparison

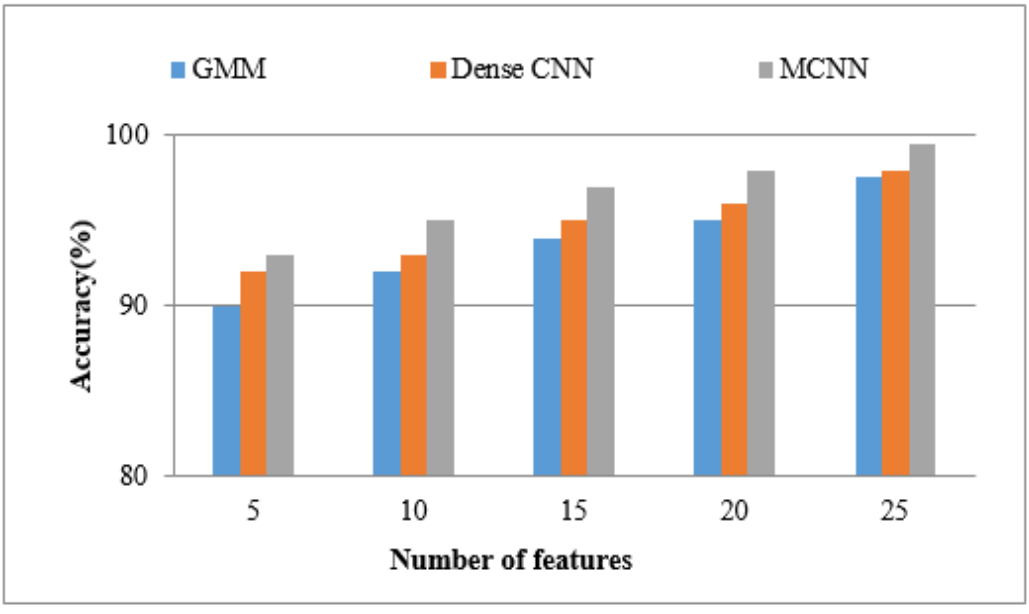

Figure 8: Result of Accuracy

Fig.8 shows the comparison of accuracy between the proposed and existing models. The MFCNN shortens processing time while improving accuracy. Since the MFCNN requires a small number of derived factors during pre-processing, it achieves an accuracy of 99.56% when compared to all other models. The BCO method was considered to segment the three regions with optimal threshold vector and hence the proposed system improved the effectiveness of detecting the position of the tip of the micropipette.

5 Conclusion

The proposed work focuses on the development of safe, reliable and autonomous positioning of the micropipette by exploring its shadow. The scope of this work can be found in the autonomous cannulation of blood vessels in the retina. The experiment pertaining to the proposed work assesses the accuracy as well as the success rate of the autonomous positioning of the micropipette. The success rate of 99.56% is achieved when the micropipette is positioned at a mean height of 1.37 mm. The proposed DL method namely the MCNN aids in the improvement of the position of the tip of micropipette tips and their shadow. Assessing the performance of this method in a clinically realistic environment is also discussed. The nature-inspired BCO-motivated segmentation as well as EFG based pre-processing adds to the accurate detection. Nevertheless, this model has a few limitations like more computation time that serves as a potential future research direction. Construction of DL models for reducing the computational time demands deciding the parameters and hyperparameters. Hyperparameter tuning is one of the attractive research scopes pertaining to the deployment of DL models in predicting the positions of pits.

6 Availability of Data and Material

Data can be made available by contacting the corresponding author.

7 Acknowledgement

The authors appreciate the support extended by Dhofar University,Salalah,Oman for providing an environment to enhance the research opportunities.

8 References

[1] Amodeo A, Linares Quevedo A, Joseph JV, Belgrano E, Patel HRH. Robotic laparoscopic surgery: cost and training. Minerva Urol Nefrol. 2009;61:121–8.

[2] Leal Ghezzi T, Campos Corleta O. 30 Years of Robotic Surgery. World J Surg. 2016;40(10):2550–7.

[3] Ma, R., Vanstrum, E. B., Lee, R., Chen, J., & Hung, A. J. (2020). Machine learning in the optimization of robotics in the operative field. *Current opinion in urology*, *30*(6), 808.

[4] De Smet MD, Naus GJL, Faridpooya K, Mura M. Robotic-assisted surgery in ophthalmology. Curr Opin Ophthalmol. 2018;29:248–53.

[5] Urias, M. G., Patel, N., He, C., Ebrahimi, A., Kim, J. W., Iordachita, I., & Gehlbach, P. L. (2019). Artificial intelligence, robotics and eye surgery: are we overfitted?. *International Journal of Retina and Vitreous*, *5*(1), 1-4.

[6] Sunshine S, Balicki M, He X, Olds K, Kang J, Gehlbach P, et al. A force-sensing microsurgical instrument that detects forces below human tactile sensation. Retina. 2013.

[7] Marban A, Srinivasan V, Samek W, Fernández J, Casals A. A recurrent convolutional neural network approach for sensorless force estimation in robotic surgery. Biomed Signal Process Control. 2019;50:134–50.

[8] Levine S, Pastor P, Krizhevsky A, Ibarz J, Quillen D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. Int J Rob Res. 2018;37(4–5):421–36.

[9] Meenink, H. C. M. (2011). Vitreo-retinal eye surgery robot: sustainable precision. Technische Universiteit Eindhoven. https://doi.org/10.6100/IR717725

[10] Tayama, T., Kurose, Y., Marinho, M. M., Koyama, Y., Harada, K., Omata, S., ... & Mitsuishi, M. (2018, July). Autonomous positioning of eye surgical robot using the tool shadow and kalman filtering. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1723-1726). IEEE.

[11] Li, P., Hou, X., Wei, L., Song, G., & Duan, X. (2018, August). Efficient and low-cost Deep-Learning based gaze estimator for surgical robot control. In *2018 IEEE International Conference on Real-time Computing and Robotics (RCAR)* (pp. 58-63). IEEE.

[12] Li, Z., Shahbazi, M., Patel, N., O'Sullivan, E., Zhang, H., Vyas, K., ... & Taylor, R. H. (2020). Hybrid robot-assisted frameworks for endomicroscopy scanning in retinal surgeries. *IEEE transactions on medical robotics and bionics*, *2*(2), 176-187.

[13] Qiu, H., Li, Z., Yang, Y., Xin, C., & Bian, G. B. (2020). Real-time iris tracking using deep regression networks for robotic ophthalmic surgery. *IEEE Access*, *8*, 50648-50658.

[14] W. Zhang, M. L. Smith, L. N. Smith and A. Farooq, "Gender and gaze gesture recognition for human-computer interaction", *Comput. Vis. Image Understand.*, vol. 149, pp. 32-50, Aug. 2016.

[15] H. Heo, W. O. Lee, J. W. Lee, K. R. Park, E. C. Lee and M. Whang, "Object recognition and selection method by gaze tracking and SURF algorithm", *Proc. Int. Conf. Multimedia Signal Process.*, vol. 1, pp. 261-265, May 2011.

[16] K. Tamura, K. Hashimoto and Y. Aoki, "Head pose-invariant eyelid and iris tracking method", *Electron. Commun. Jpn.*, vol. 99, no. 2, pp. 19-27, 2016.

[17] B. Li and H. Fu, "Real time eye detector with cascaded convolutional neural networks", *Appl. Comput. Intell. Soft Comput.*, vol. 2018, pp. 1-8, Apr. 2018.

[18] H. Kannan, "Eye tracking for the iPhone using deep learning", 2017.

[19] F. Wolfgang, S. Thiago, K. Gjergji and K. Enkelejda, "PupilNet: Convolutional neural networks for robust pupil detection", *Revista De Odontologia Da Unesp*, vol. 19, no. 1, pp. 806-821, 2016.

[20] S. Hoffman, R. Sharma and A. Ross, "Convolutional neural networks for iris presentation attack detection: Toward cross-dataset and cross-sensor generalization", *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, pp. 1620-1628, Jun. 2018.

[21] He, C., Patel, N., Shahbazi, M., Yang, Y., Gehlbach, P., Kobilarov, M., & Iordachita, I. (2019). Toward safe retinal microsurgery: Development and evaluation of an rnn-based active interventional control framework. *IEEE Transactions on Biomedical Engineering*, *67*(4), 966-977.

[22] Keller, B., Draelos, M., Zhou, K., Qian, R., Kuo, A. N., Konidaris, G., & Izatt, J. A. (2020). Optical coherence tomography-guided robotic ophthalmic microsurgery via reinforcement learning from demonstration. *IEEE Transactions on Robotics*, *36*(4), 1207-1218.

[23] Zhou, M., Wang, X., Weiss, J., Eslami, A., Huang, K., Maier, M., ... & Nasseri, M. A. (2019, May). Needle localization for robot-assisted subretinal injection based on deep learning. In *2019 International Conference on Robotics and Automation (ICRA)* (pp. 8727-8732). IEEE.

[24] Hutchinson S, Hager GD, Corke P. A Tutorial on Visual Servo Control.IEEE Trans Rob Autom; 1996: vol. 12, no. 5, p. 651-670.

Mr. Mukesh Madanan is a Senior Lecturer of Computer Science at Dhofar University, Oman. He received his B.Tech in Computer Science and Engineering from M. G. University, India and went on to complete his MSc.in Software Engineering from the University of Portsmouth, UK. He is currently pursuing PhD in Information & Communication Technology from UNITEN, Malaysia. His areas of research include Machine Learning, Deep Learning, Robotics, Software Methodologies, IoT and Computer Networks

Dr. Nurul Akhmal Mohd Zulkefli is an Assistant Professor at the Department of Computer Science, College of Arts & Applied Sciences, Dhofar University, Oman. She received her Ph.D. in Information Technology (Expert Decision System) from Universiti Teknologi PETRONAS, Malaysia. Her research interests are in Expert Decision Systems (Multi-Criteria Decision Making), Big Data, Knowledge Management Systems, Software Engineering, and IT in Education.