



An Improved Fuzzy Deep Learning (IFDL) model for managing uncertainty in classification of pap-smear cell images

Mona Benhari, Rahil Hosseini*

Department of Computer engineering, Shahr-e-Qods Branch, Islamic Azad University, Tehran, Iran



ARTICLE INFO

Keywords:
 Deep learning
 Uncertainty handling
 Dempster-Shafer theory of evidence
 Belief network
 Cell image classification, Fuzzy systems

ABSTRACT

Applications of deep learning models for medical image analysis have been concentrated in the recent years. An automatic detection system to detect the class of Pap smear cell and cervical cancer is a challenging problem due to time consuming and erroneous process of the detection for technicians. This study presents an improved Deep Convolutional Neural Network (DCNN) for analysis of Pap smear images for early detection of cervical cancer. The proposed model addresses the issue of classification of samples with similar probability in classification layer of a DCNN. To address this challenge, an Improved Fuzzy Deep learning (IFDL) model has been proposed by taking advantages of Deep Belief Network using Dempster combination rule, and Fuzzy weighting system, to manage uncertainty of similar classes in the classification layer. In this method, a new layer by Belief Networks using Dempster combinational rule, aggregates the evidences to handle uncertainty of assigning correct class, between different classes. To address the issue of the object rejection in Belief network, a fuzzy weighting system has been proposed. The experimental results for two classes problem and seven classes problem on Herlev cell image dataset, show the superiority of the proposed model. This model with an accuracy of 99.20% outperforms counterpart methods and is promising for early detection of the cervical cancer.

1. Introduction

Cervical cancer is one of the cancerous diseases with high mortality rate across the world, especially in developing countries. Pap smear test is an annual test recommended to the woman for early detection or prevention of cervical cancer. In Pap smear test thousands of cervical cells are analyzed by a cytologist to detect abnormal cells. Pap smear screening by cytologists is a tedious and time-consuming process, therefore using an automatic Pap smear screening system deems essential.

Pap smear images are classified into seven types: normal superficial, normal intermediate, normal columnar (which are related to the different layers of cervical tissue), and light dysplastic, moderate dysplastic, severe dysplastic, and carcinoma (which show malignancy of a cell).

The main challenge in Pap smear cells detection is that the columnar cells are very similar to dysplastic cell, due to the size of cell nuclei and the cell appearance. This issue causing high degree of uncertainty in the classification process. Therefore, uncertainty handling is important to ameliorate the reliability of a classification algorithm to get close to a

desired human behavior (without carrying any pre-judgment or common human errors indeed). The possible sources of uncertainty in a medical system diagnosis like Pap smear image analysis is listed below (Hosseini et al., 2010):

Imprecision in input data and features of patterns in microscopic Pap smear images and noisy measurements, Intra-uncertainty of an expert and the hesitation to classify the cells, The word perception of an expert by the others, The non-stationary features, Uncertainty of mathematical models for measuring the complex features of images, Inter uncertainty and the uncertainty of the classification of the cells between experts, and Uncertainty of all processes of image enhancement, segmentation, feature extraction and, classification methods.

More accurate image enhancement and segmentation leads to more accurate classification. Moreover, the feature design and feature extraction plays an important role in classification process. Each error in each phase transfers to next phase and decreases the performance of the system. A novel approach to combining the process of image enhancement, segmentation, feature extraction and automatic feature extraction to improve classification performance is the Deep Convolutional Neural Network, which has achieved promising results in recent

* Corresponding author.

E-mail address: rahil.hosseini@qodsiau.ac.ir (R. Hosseini).

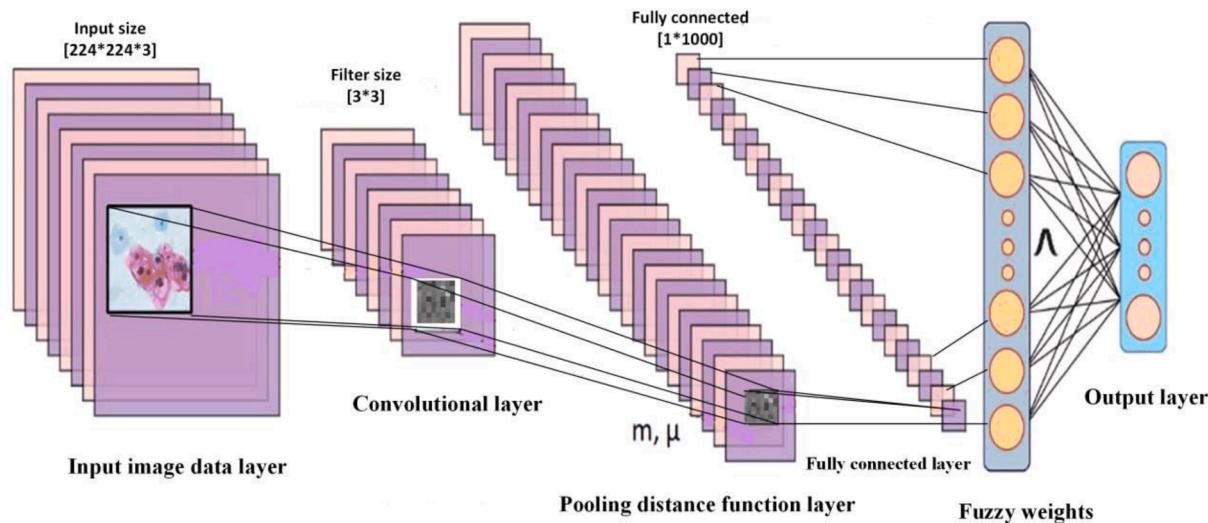


Fig. 1. Architecture of the proposed IFDL model.

studies. However, DCNN studies did not manage the uncertainty of the classification process, precisely in classifying similar cells with different types. This issue is addressed in this study. The proposed IFDL model can manage uncertainty to address the ambiguity associated with similar cells' appearance in the classification process.

For the first time Zheng Tong et al (Tong et al., 2019), proposed an object detection method using Deep CNN and Dempster-Shafer (DS) theory to manage uncertainty of the classification process. In this method the evidences combine together with DS theory to compute each class probability. The higher-class probability is the winner class for the input object. If the probability of the winner class is less than a threshold the object is rejected. But this rejection may ignore the valuable data in the dataset and this method is not proper for limited medical images.

This study proposed a novel Improved Fuzzy Deep Learning (IFDL) model, which takes advantage of the theory of the Bayes Belief network for cell image classification. Furthermore, fuzzy logic has been applied to address the uncertainty associated with pap smear sample rejection (unclassified samples) in the output of the Bayes Belief network. The novel IFDL model can manage uncertainty in the classification process and address the issue of rejected samples, which is very valuable, especially in classifying Pap smear cells in the Bayes Belief Network. The proposed IFDL model is beneficial for medical image processing with low volumes of image datasets and uncertainty in classifying abnormalities.

2. Literature review

Traditional Computer Aided Detection (CAD) systems or every image classification system (Dash et al., 2022; Woźniak et al. 2021) have four phase of image enhancement, segmentation, feature extraction and classification. In image enhancement and image segmentation has been investigated in Sahba et al. (2003), Lu et al. (2015), Nosrati and Hamarneh (2015), Hansang Lee and Junmo Kim (2015), Lu et al. (2015). Image enhancement and image segmentation in a computer aided detection system (CAD) plays an important role. The more accurate segmentation the more accurate detection system. Due to this importance several segmentation challenges specially cell segmentation has been held in 2014 and 2015, (Masoud S. Nosrati and Ghassan Hamarneh, 2105; Hansang Lee Junmo Kim, 2015; Araujoa et al., 2018). For handling segmentation uncertainty (William et al., 2019) proposed a Weka segmentation and a sequential approach for image enhancement. In classification systems specially learning process, the proper number of features to avoid overfitting are very important. For feature extraction and feature selection (William et al., 2019), applied evolutionary

algorithm and Mohammad Subhi et al. (Al-Batah et al., 2014) Proposed an Automatic Feature Extraction (AFE) algorithm for feature selection and extraction. But the errors of these steps transfer to the classification step. To address this challenge and to have more accurate detection Convolutional Neural Networks (CNN) has been applied in recent studies (Cheng et al., 2022; Tang et al., 2022; Huang et al., 2021). The CNN extracts the Feature maps through the layers from input images without any separate segmentation process. Moreover, as it was mentioned in previous section using (DCNN), because of direct classification from the input cell images without dependency to the segmentation and feature selection steps, leaded to more accurate classification and cell detection (Martinez et al., 2020; Wu et al., 2019; Abhinaav, and Brindha, 2019; Bora et al., 2016) with more than 90% accuracy. But these studies faced with limitations in cell image sources, especially in Pap smear images, therefore the results are not reliable. To solve this problem pre trained CNN and transfer learning method has been applied (Harangi et al., 2019; Twinkle Dalal and Manjeet Singh, 2021; Long D. Nguyen et al., 2018; Dharani et al., 2020; Sompawong et al., 2019; Taha et al., 2017; Zhang et al. 2018; IbrahimWaly et al., 2021; William et al., 2018; William et al. 2019)which approved promising results more than 95%. Moreover, many studies in last 2 years, (OrhanYaman and TurkerTuncer, 2022; Khamparia et al., 2021; Kumararaja & Deepa, 2021; Mohammed et al., 2021; Diniz et al. 2021; Bhatt et al., 2021; Palanisamy et al., 2021; Tripathi et al., 2021; Singh et al., 2021; Rahaman et al., 2021; Basak et al., 2021) increased classification accuracy more than 98%. But none of these studies consider uncertainty handling.

In the Proposed IFDL model the uncertainty of classification process have been addressed.

The rest of this study is organized as follows: Section 2, explains the methods and materials in the form of proposed model. Section 3 includes the theory of the model. Section 4 includes the discussion of the experimental results. Finally, the research findings are concluded in Section five.

3. Methods and materials

in this section the architecture of the method and the steps of the model are explained in detail.

3.1. The architecture of proposed IFDL model

This section investigates the mathematical model and architecture of the proposed model. The suggested model, presents a new CNN archi-

ture with the capability of uncertainty handling in a deep structure using Dempster combinational rule and fuzzy set theory. In this study, the Belief networks use Dempster-Shafer rule for evidence combination. The independent evidences are aggregates with DS theory to handle uncertainty in classification method. Also, fuzzy logic has been combined to the Belief network to improve the efficiency of the system and prevent the object rejection of the classes with low probability. The model inputs are cell images and the class of the cell images are considered as the output of the system. Fig. 1 shows the architecture of the proposed model. The Pap smear cell images, are the inputs of convolutional layer and the features are extracted through the deep layers. The extracted feature maps in fully connected layer are sent to distance layer for calculation of the Euclidian distance between each feature map and the prototypes (the prototypes are the feature map of the representative sample in dataset). The Belief Function layer computes the probability of each class as vector m and μ , which μ is the copy of m that are combined with vector m in the next layer. The Dempster rule combine the m and μ . The fuzzification layer for weight adjustment is suggested to handle cell overlapping uncertainty and was applied to the DS layer for better prediction.

In the proposed IFDL model, the weights of the feature maps are updated during the convolutional layers. The pooling distance function layers calculate the distance between each feature map and each prototype. The prototypes show the representative of the dataset, which is selected by k-means clustering. The output nodes of the distance function layer are the results of the distance between a node and each prototype. These values are considered events and are aggregated with the Dempster rule. In order to address the issue of sample rejection (unclassified samples) in the output of the Belief network then, the fuzzy logic process was applied to make the decision about the class of the samples.

The input tensor after augmentation is [224 224 3], and the last convolution operation tensor before the fully connected layer in Google-Net is an inception layer filter size is [3 3]. Number of channels are 832, number of filters are 128, stride is [1 1], and padding size is [0 0 0]. Also, the fully connected layer has the size of [1 1000].

The proposed IFDL model includes the following layers:

Convolutional layers, the convolutional layers extract the feature maps of the input image data. The features are sent to Belief network to calculate the distances (Eq. 6).

Apply Belief function, the feature maps enter to the Belief network. The Euclidian distance of each feature map and prototypes (The selected representative of the dataset) are calculated, then the probability of each class is computed (Eq. 7).

Dempster combinational rule: the probability of different classes is aggregated using Dempster rule (Eq. 9).

Fuzzification process: In order to prevent rejection of the Belief network, a fuzzy model of the feature space has been applied to the Dempster rule. This fuzzy system method uses all of the image cells, and prevent the sample rejection of the Belief network.

The following subsection explains the basic theory of the proposed model in details.

4. Theory of the Belief network using Dempster rule

In this section a brief overview of the Belief network and Dempster rule as the base theory, is presented (Tong et al., 2019). One of the efficient tools for uncertainty modeling in decision making systems is the theory of evidence, presented by Shafer in 1976 (G., 1976) and completed by Dempster. Denoeux (Tong et al., 2019); Denoeux, 2000 applied this theory in Belief Networks and also in the CNN.

Assumed $\Omega = \{w_1, w_2, \dots, w_q\}$ as a finite subset of different classes, a mass function m () is a function from 2^Ω to $[0, 1]$ with the following conditions:

$$m(\phi) = 0 \quad (1)$$

$$\sum_{A \in \Omega} m(A) = 1 \quad (2)$$

$m(A)$ is the amount of belief we have in A, for any $A \subset \Omega$. Set A is called a focal element of m when $m(A) > 0$.

Credibility function $bel(A)$ is the lower bound of uncertainty and a Plausibility function $pl(A)$ is the upper bound of uncertainty, related to m for any $A \subset \Omega$, are defined as,

$$bel(A) = \sum_{B \subset A} m(B) \quad (3)$$

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (4)$$

The quantity $bel(A)$ is interpreted as a global measure of one's belief that hypothesis A is true, while $pl(A)$ is the amount of belief that could potentially be placed in A.

Dempster advised a rule called Dempster Rule shown by \oplus operator which can combine two mass functions m_1 and m_2 . These mass functions represent the independent items of evidence, as follows:

$$m(\emptyset) = 0, (m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{\sum_{B \cap C \neq \emptyset} m_1(B)m_2(C)} \quad (5)$$

In this rule, if $A \neq \emptyset$ and $(m_1 \oplus m_2)(\emptyset) = 0$, all masses can be combined (Denoeux, 2000).

Belief function provides an adaptive pattern classifier based on Dempster rule. Each class has some patterns or prototypes considered as evidence related to their classes. For each evidence method, a mass function is assigned which is combined by Dempster rule. The Belief activation function's Eq.s are explained in Denoeux (2000).

The set $\Omega = \{w_1, w_2, \dots, w_q\}$ are the different classes in Pap smear images, since this classification method is two class problem (normal and abnormal), the set has 2 members. Each cell is the evidence in Belief theory and each prototype selected from image dataset, and then m is computed.

The input $x \in \mathbb{R}^P$ is entered into the model for the classification into one of the M classes, (in our dataset M=2 normal and abnormal). The classes are shown by w_1, \dots, w_m and T is the training set of N samples with P-dimension. The uncertainty of the class x is calculated by mass function. This mass function or weight function is calculated for each class's weight $\Omega = \{w_1, \dots, w_M\}$. The algorithm has five steps in the form of layers of IFDL. The n prototypes p^1, \dots, p^n are the weight vector of the distance function layer, the dimensions of vector p is $k \times s$, which k extracted from k-means clustering algorithm to find the best representative of the samples. k is selected heuristically, and s is equal to input sample size. It should be mentioned that the samples are the feature maps of the fully connected layer as the input of the classification layer, and prototypes are the representative of the samples selected by k-means algorithm.

Step1: The distance function layer calculates the distance between each input and each prototype.

$$d^i = \|x - p^i\| \quad i = 1, \dots, n \quad (6)$$

Step2: For each d^i , an activation function with the parameters $\eta^i (\eta^i \in \mathbb{R})$, and $\alpha^i \in (0, 1)$ is as,

$$s^i = \alpha^i \exp(-(\eta^i d^i)^2) \quad (7)$$

Step3: The mass function m^i for each prototype p^i is calculated as follows:

$$m^i = (m^i(w_1), \dots, m^i(w_i), m^i(\Omega))^T = (u_1^i s^i, \dots, u_M^i s^i, 1 - s^i)^T \quad (8)$$

Where $u^i = (u_1^i, \dots, u_M^i)$ is the prior probability related to each prototype p^i that $\sum_{j=1}^M u_j^i = 1$.

Step4: The node weights are calculated by the same mass functions

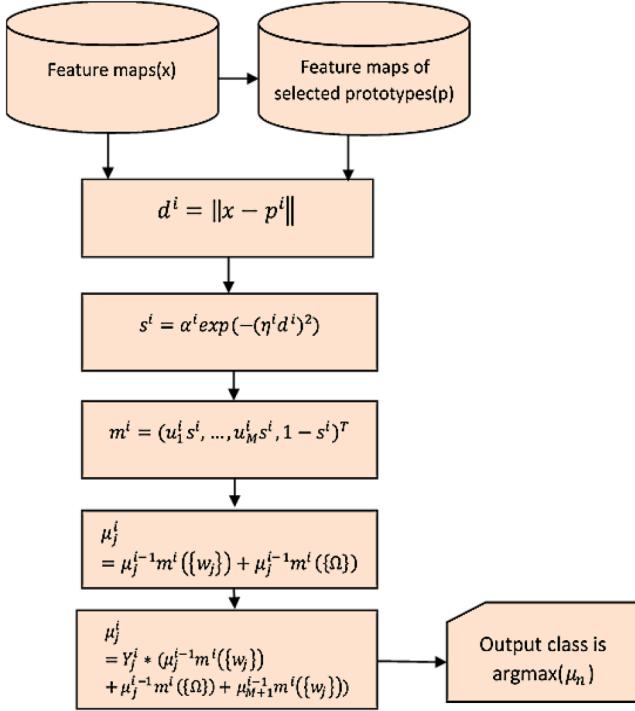


Fig. 2. Steps of the IFDL model block diagram.

m^i , $i = 1, \dots, n$, are combined with Dempster rule. The $\mu_i = (\mu_1^i, \dots, \mu_1^i)$, $i = 1, \dots, n$ which is calculated by m is defined by the following Eq.:

$$\begin{aligned} \mu^1 &= m^1 \\ \mu_j^i &= \mu_j^{i-1} m^i(\{w_j\}) + \mu_j^{i-1} m^i(\{\Omega\}) + \mu_{M+1}^{i-1} m^i(\{w_j\}) \end{aligned} \quad (9)$$

For $i = 1, \dots, n$ and $j = 1, \dots, M$ and $\mu_{M+1}^i = \mu_{M+1}^{i-1} m^i(\{\Omega\})$ $i = 2, \dots, n$

Eq. 9 defines Dempster orthogonal sum. In this Equation the combination of evidences is performed by three sums of multiplying m and μ in different nodes. Denoeux (Denoeux, 2000) represents the classification by last node of μ . For each class 1 to $M+1$, the max probability is the winner class and if the probability of $M+1$ th element is lower than specific threshold, the sample is rejected. To prevent this sample rejection (Denoeux, 2000), the output of the Dempster combination rule is sent to the proposed fuzzy layer and final decision about the sample class is made. Fuzzy logic has high potential for managing uncertainty and provides high interpretability in decision support systems. A fuzzy set contains a class of objects with a grade of membership. The membership grade is assigned by a membership function which is a number between zero and one. This assignment is proper for approximate reasoning. More often in real world the criteria of membership functions are not precise. For example, in Pap smear image classification some cells have an ambiguous assignment to a specific class. Fuzzy sets are capable of modelling overlapping classes in the Pap smear cell types. One of the major challenges and vagueness in the Pap smear cell image analysis and classification, is that the normal columnar cells seem similar to the cancerous cells regarding their shapes and sizes, leading to sample rejection in response of the Dempster combination rule. To manage this sort of uncertainty, the capability of fuzzy systems has been combined with the Belief theory for the classification of Pap smear cells. To design such a fuzzy inference system, the features extracted by the convolutional layer are used as the input of the fuzzy system. The extracted features are considered as input vector F including n samples, and k features, as:

$$F_j^i = \left(f_j^i, \dots, f_n^i \right)^T \quad (10)$$

The output of the fuzzy system is defined as a weighting factor used

to combine with the output of Dempster rule to prevent sample rejection. This fuzzy system improves classification efficiency by considering overlaps in cell classes. The fuzzy output is defined by Eq. 11. The output number of the fuzzy system is equal to M classes. The output value of the fuzzy system is assigned to the nearest class and its complement is assigned to the other class.

$$Y_j^i = (y_1^i, \dots, y_M^i)^T \quad (11)$$

Step5: Eq. 12 defines how the fuzzy system is incorporated in Dempster rule.

$$\mu_j^i = Y_j^i * \left(\mu_j^{i-1} m^i(\{w_j\}) + \mu_j^{i-1} m^i(\{\Omega\}) + \mu_{M+1}^{i-1} m^i(\{w_j\}) \right) \quad (12)$$

Belief network classifier using Dempster Shafer theory performs prediction with $M+1$ parts, M parts related to the M classes, and the $M+1$ th part related to sample rejection. This rejection may lead to the loss of valuable information especially when the numbers of samples are limited. To prevent sample rejection and improve classification efficiency for handling the high degree of uncertainty in the classification method, the response of the designed fuzzy weighting system is used. Therefore, the fuzzy system predicts the class of the network output in the case of sample rejection. the steps of the method are summarized in a block diagram by Fig. 2.

4.1. Learning process of the Belief network

The parameters used in Eqs. (7) and (8), α^i , η^i , u^i , are initialized as the followings:

$$\gamma^i = (\eta^i)^2 \quad (13)$$

$$\alpha^i = \frac{1}{1 + \exp(-\xi^i)} \quad (14)$$

$$u_j^i = \frac{(\rho_j^i)^2}{\sum_{k=1}^M (\rho_j^k)^2} \quad (15)$$

To train the network, these parameters are updated by partial differential of the IFDL error. The network error in neural network application is usually obtained by the difference between the predicted class and the actual class. For M Classes $M+1$, the probability is calculated by the network but the actual has M part. The $M+1$ th part is applied to the M part with a fraction of v . If v is equal to 0, the credibility, if v is equal to $1/M$ the plausibility, and if v is equal to 1, the middle probability of the class w_j is calculated. The target vector is defined by: $t = (t_1, t_2, \dots, t_m)$ and the predicted vector is defined by $p_v = (p_{v,1}, p_{v,2}, \dots, p_{v,M})$ when $p_{v,q} = m_q + v m_{M+1}$. With the definition of target vector and prediction vector, the error is defined as (Denoeux, 2000):

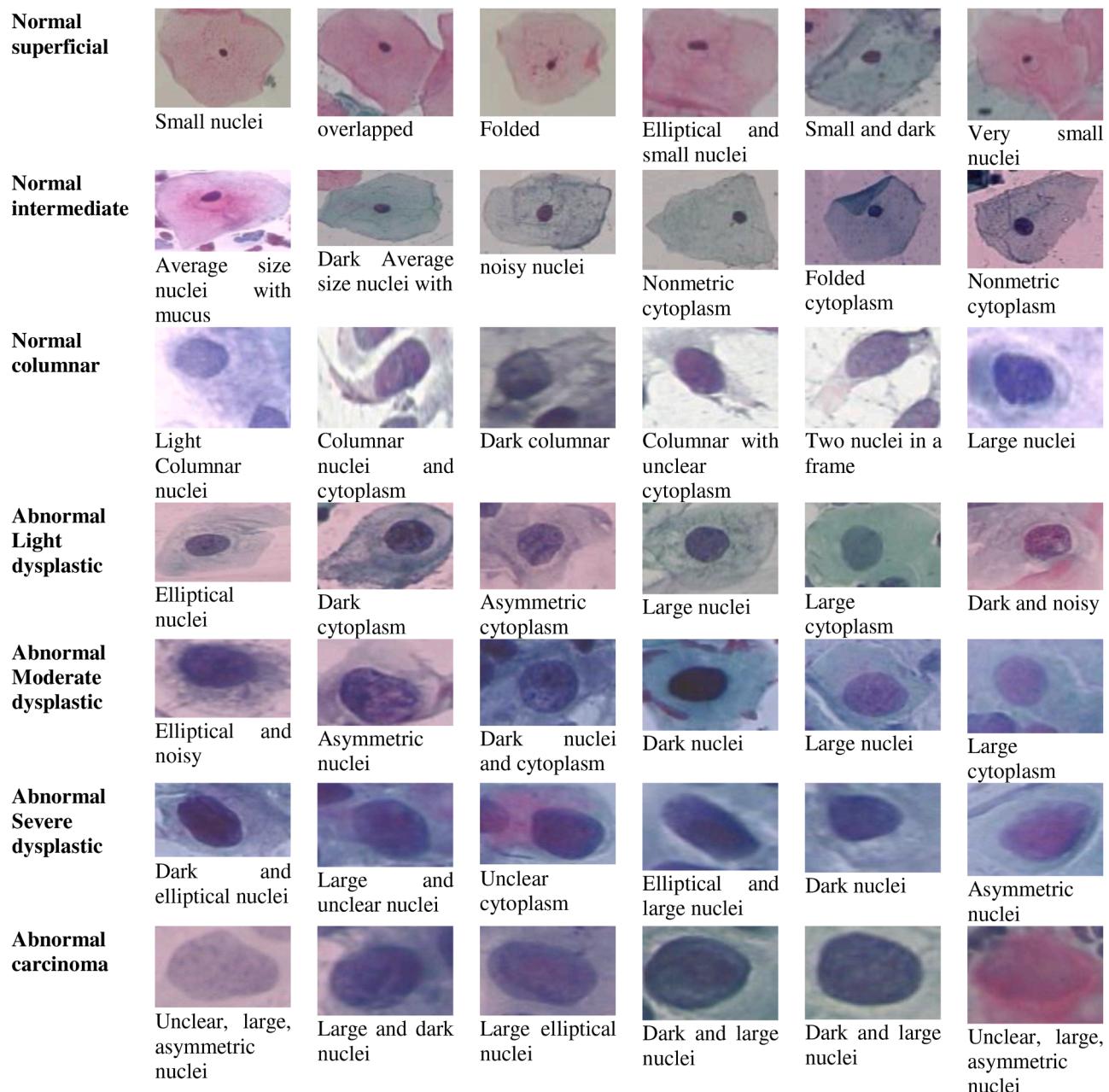
$$E_v = \frac{1}{2} \sum_{q=1}^M (p_{v,q} - t_q)^2 \quad (16)$$

In the learning process, the error of the network in each iteration is defined by the mean error of the training set X with size N :

$$E_v = \frac{1}{N} \sum_{x \in X} E_v(x) \quad (17)$$

Minimizing the E_v in learning process is explained in the following.

Learning process contains the optimization parameter to minimize the error function in Eq. 16. First the parameter initialization with the constraints of $0 < \alpha < 1$, $\gamma > 0$, $\sum_{j=1}^M u_j^i = 1$ is performed, which i shows the number of prototypes and j is the number of classes, then in each iteration for each training sample error is calculated by 16 and the mean of the training error is calculated according to Eq. 17. The parameters are updated based on Gradient optimization using Eqs. 18 to 25.

**Fig. 3.** Dataset cell types.

The partial differential of E related to parameter, s^i , $i = 1, \dots, n$, where n is the number of prototypes, equals to:

$$\frac{\partial E_v(x)}{\partial s^i} = \sum_{j=1}^M (p_{v,j} - t_j) \left(u_j^i (\bar{m}_j^i + \bar{m}_{M+1}^i) - \bar{m}_j^i - v \bar{m}_{M+1}^i \right) \quad (18)$$

While \bar{m}_j^i obtained by mass function output as the conjunctive combination of two masses of m :

$$\bar{m}_j^i = \frac{m_j - \left(\frac{m_{M+1} \times m_j^i}{m_{M+1}^i} \right)}{m_j^i + m_{M+1}^i} \quad j = 1, 2, \dots, M \quad (19)$$

$$\bar{m}_{M+1}^i = \frac{m_{M+1}}{m_{M+1}^i} \quad (20)$$

From the m complement Equation, the $\partial m / \partial u$ will be:

$$\frac{\partial m}{\partial u_j^i} = s^i \left(\bar{m}_j^i + \bar{m}_{M+1}^i \right) \quad (21)$$

According to Eq. (12), u_j^i is related to β_j^i . Therefore the derivative E_v to β_j^i is performed:

$$\frac{\partial E_v(x)}{\partial \beta_j^i} = \frac{2\beta_j^i}{\left(\sum_{l=1}^M (\beta_l^i)^2 \right)} \left[\frac{\partial E_v(x)}{\partial u_j^i} \sum_{l=1}^M (\beta_l^i)^2 - \sum_{l=1}^M (\beta_l^i)^2 \frac{\partial E_v(x)}{\partial u_l^i} \right] \quad (22)$$

l is the index of $k = 1, 2, \dots, m$ when $k \neq j$.

Additionally, there are two more parameters that construct α^i and γ^i in Eqs. (13) and (14). These parameters ξ^i and η^i calculate the error derivatives:

$$\frac{\partial E_v(x)}{\partial \xi^i} = \frac{\partial E_v(x)}{\partial s^i} \exp \left(-(\eta^i d^i)^2 \right) (1 - \alpha^i) \alpha^i \quad (23)$$

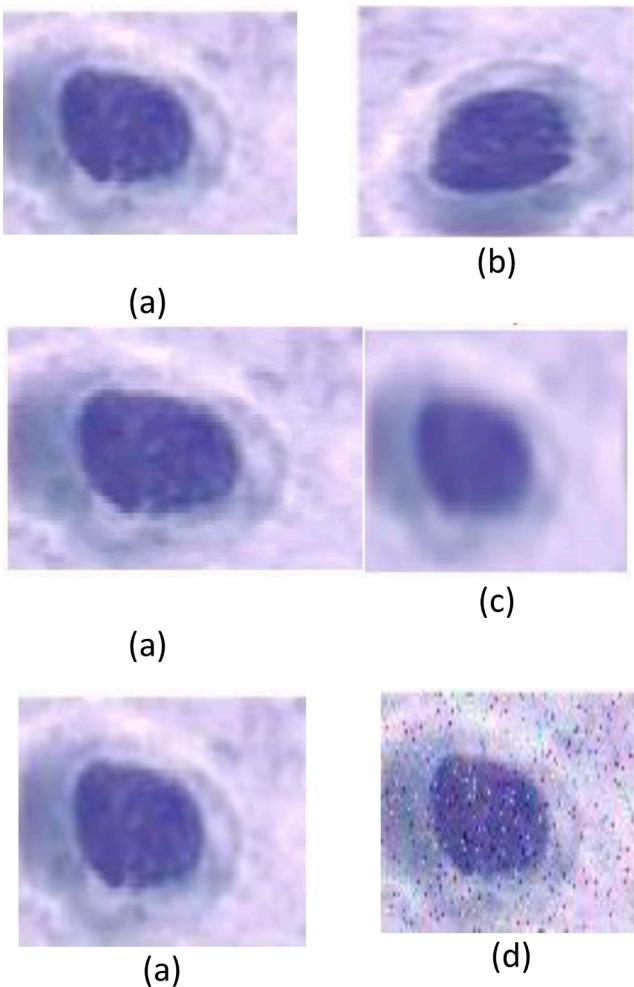


Fig. 4. Data augmentation process.

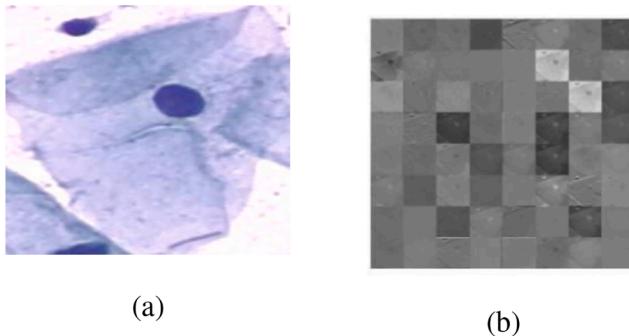


Fig. 5. the extracted features in the third convolution layer.

Table 1
Two class problem results of the IFDL model.

Fold	Accuracy%	Specificity%	Sensitivity%	AUC%
1	98.85	98.92	98.78	99.94
2	98.85	100	97.84	98.96
3	98.86	98.61	99.03	98.94
4	98.28	98.76	98.87	99.97
5	97.71	97.75	97.77	99.92
Average	98.51	98.80	98.45	99.95

Table 2
Seven class results.

Class #	Specificity%	Sensitivity%	AUC%
Class 1	83.51	72.38	95.74
Class 2	75.33	88.97	94.76
Class 3	77.38	63.72	92.19
Class 4	83.72	78.83	96.34
Class 5	93.87	93.26	99.71
Class 6	95.09	93.87	99.23
Class 7	68.58	78.10	93.15

Table 3
The results of proposed IFDL model for 2-class.

Method	Accuracy%	Specificity%	Sensitivity%	AUC%
IFDL with Google-Net	99.20	99.75	99.25	99.97
IFDL with Res-Net50	99.20	99.25	98.51	99.99
IFDL with Alex-Net	98.80	99.78	99.25	99.91

Table 4
Comparisons of the IFDL to base methods.

Method	Accuracy%	Specificity%	Sensitivity%	AUC%
IFDL	98.51	98.80	98.45	99.95
CNN	90.71	47.05	75.90	73.29
BF	93.39	95.04	85.13	96.16
FL	89.10	86.93	71.03	80.12

$$\frac{\partial E_v(x)}{\partial \eta^i} = \frac{\partial E_v(x)}{\partial s^i} \frac{\partial s^i}{\partial \eta^i} = \frac{\partial E_v(x)}{\partial s^i} \left(-2\eta^i (d^i)^2 s^i \right) \quad (24)$$

Finally, to update the prototype the error derivatives related to p^i are:

$$\frac{\partial E_v(x)}{\partial p^i} = \frac{\partial E_v(x)}{\partial s^i} \frac{\partial s^i}{\partial p^i} = \frac{\partial E_v(x)}{\partial s^i} \left(-2(\eta^i)^2 s^i (x_j - p_j^i) \right) \quad (25)$$

Parameter optimization with the above Equation is preformed to minimize the error. The optimization process continues until an error threshold is satisfied or the number of iterations reaches to a specific number (Denoeux, 2000). In the next section, the experimental results are discussed.

5. Discussion

The proposed model was evaluated and its performance is reported in this section. The standard public cell images of the Herlev dataset were used to evaluate the performance of proposed model. The performance of the method was assessed by using an ROC curve analysis and the results are compared with the state-of-the-art methods. For validation purpose, average of 5-fold cross validation was applied. The equations of Accuracy, Specificity, Sensitivity, according to the true positive (TP), the samples with the true malignancy prediction, the true negative (TN), the samples with the true normal cell prediction, the false negative (FP), the samples with the false malignancy prediction, and the (FN), the samples with the false normal prediction are as bellow:

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN} \quad (27)$$

$$SPECIFICITY = \frac{TN}{FP + TN} \quad (28)$$

$$SENSITIVITY = \frac{TP}{TP + FN} \quad (29)$$

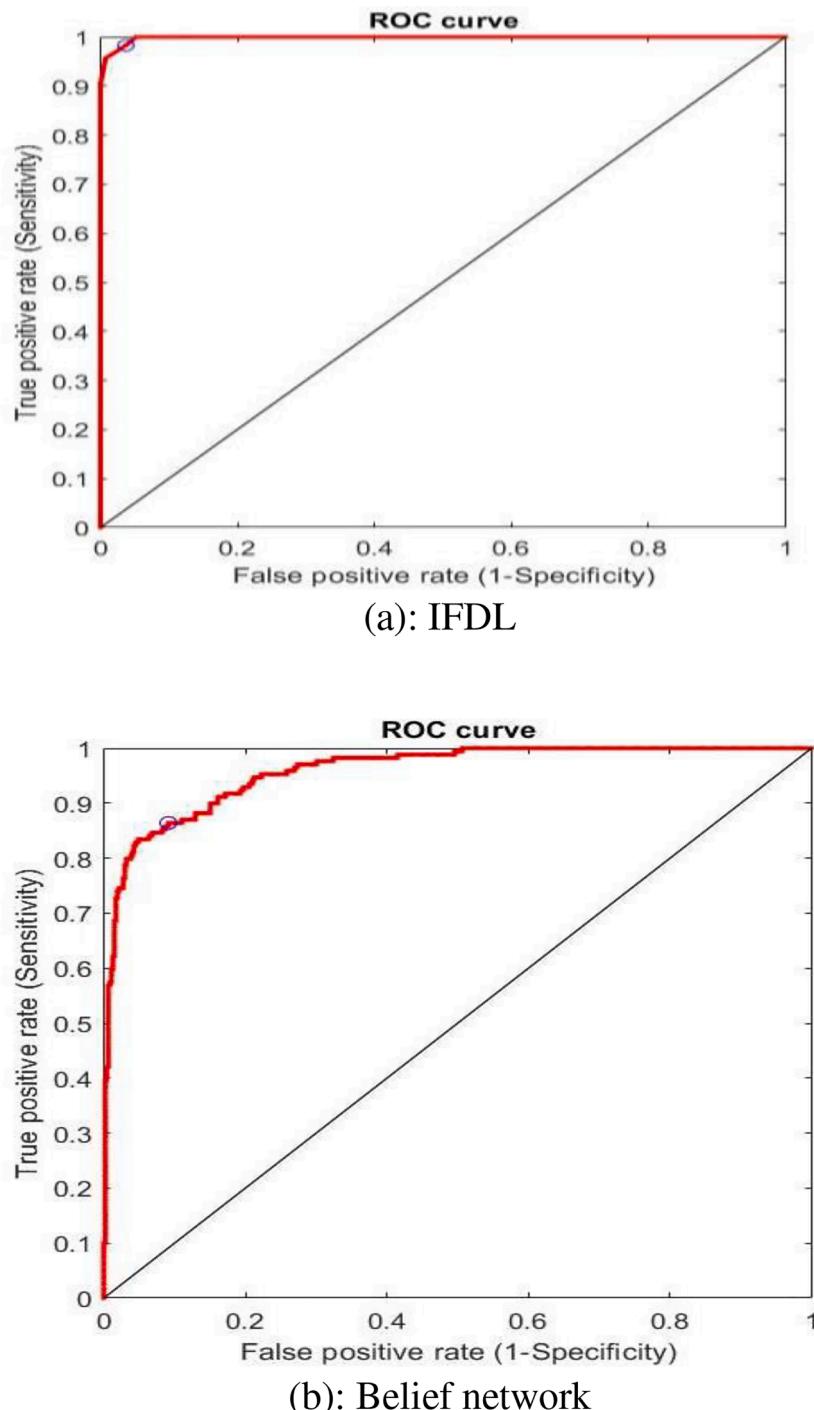


Fig. 6. The AUC of the ROC curve.

5.1. Pap smear cell image dataset characteristics

The Pap-smear cell image dataset is a well-known publicly available dataset (<http://mdeLab.aegean.gr/downloads/>) (0), was collected at the Herlev University Hospital by a digital camera and microscope. The image resolution was 0.201m per pixel. The specimens were prepared via conventional Pap smear and Pap staining. There is a total of seven different classes diagnosed by two cyto-technicians and a doctor, in order to maximize certainty of the diagnosis. These labeled cell images were used as the input of the proposed model. Pap smear images are classified into seven types: normal superficial, normal intermediate, normal columnar (which are related to the different layers of cervical

tissue), and light dysplastic, moderate dysplastic, severe dysplastic, and carcinoma (which show malignancy of a cell). Fig. 3 shows the different appearance of each cell type. It displays six samples in each class to show the diversity of the appearance in shape, size and brightness of each cell type.

There are three scenarios to implement the IFDL model on the dataset. The first scenario is to classify the Herlev dataset into normal and abnormal cells. The second scenario is to classify the Herlev dataset into seven classes described in the previous paragraph. Finally, the third scenario is to classify the SIPaKMeD (Plissiti et al., 2018) dataset into three classes. SIPaKMeD database consists of 4049 images of isolated cells of Pap smear slides. These slides include three categories of cells:

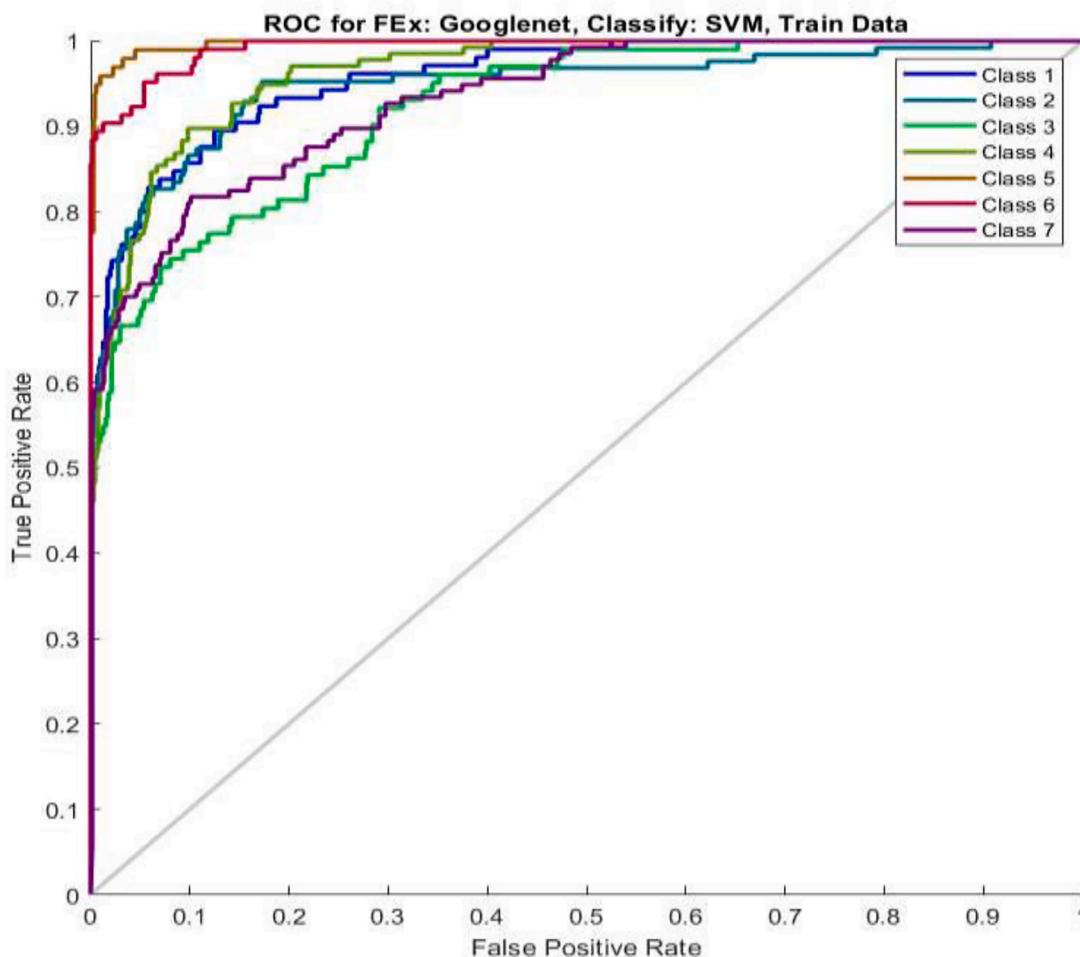


Fig. 7. The AUC of the ROC curve analysis for 7-class problem.

Table 5
The different class description.

Class #	Description
1	Normal superficial
2	Normal intermediate
3	Normal columnar
4	Abnormal Light dysplastic
5	Abnormal Moderate dysplastic
6	Abnormal Severe dysplastic
7	Abnormal carcinoma

normal (superficial and Parabasal cells), abnormal (Koilocytotic and Dyskeratotic cells), and Benign cells (Metaplastic cells).

To prevent overfitting and also to improve accuracy, image enhancement especially data augmentation plays an important role. Therefore, the data augmentation contains data rotation and image filtering has been applied to the cell images. In Pap smear image analysis especially the Herlev dataset the number of abnormal cells is much more than normal cells and it is 675 to 242, and this lack of data in normal class, leads to poor learning rate in this class. So, data augmentation for normal cells seems to be essential to have more accurate classification. For this reason, the normal cell images are augmented three times more in three steps. In Step one the normal cell images was rotated 90 degrees and add to the dataset. Step 2 the mean filter with the filter size of 3×3 was added to the dataset. Third step the salt and pepper noise were applied to the image and add to the dataset. Fig. 4 (a), shows the original cell image. In image (b), the cell was rotated 90 degrees, in image (c), a

mean filter with the size of 3×3 was applied to it, and in image (d) the salt and pepper noise was added to the cell image. The size of the dataset after this augmentation process increased to about three times more than the original dataset and augmented from 242 images to 576 (including 675 abnormal cells). The process of image augmentation increases accuracy in classification and helps the Deep Convolutional Neural Network learn both classes as well.

In order to manage the noisy images during the augmentation process the salt and pepper noises have been added to cell images to examine the robustness of the proposed IFDL method.

After data augmentation, augmented dataset was sent to the convolutional and pooling layers. Fig. 5 shows the features extracted from second convolutional layer. These feature maps are the input of the fully connected layer. Fully connected layer is the input of the classification layer Finally, the classification layer predicts the cell classes.

Next section discussed about the experimental results of the proposed IFDL model.

5.2. The experimental results

The experimental results with 5-fold cross validation for two-class problem and seven class problem is presented in Tables 1 and 2. In Table 1 The average of 5-fold cross validation with 95% confidence interval and variance of 0.26, in terms of accuracy, specificity, sensitivity, and AUC are 98.51%, 98.80%, 98.45 and 98.95%, respectively. The different fold reveals different results due to randomness selection of the train and test dataset partitioning.

In the Table 2 the specificity, sensitivity and AUC of the ROC curve of

Confusion matrix of 2-class problem



Confusion matrix of 7-class problem

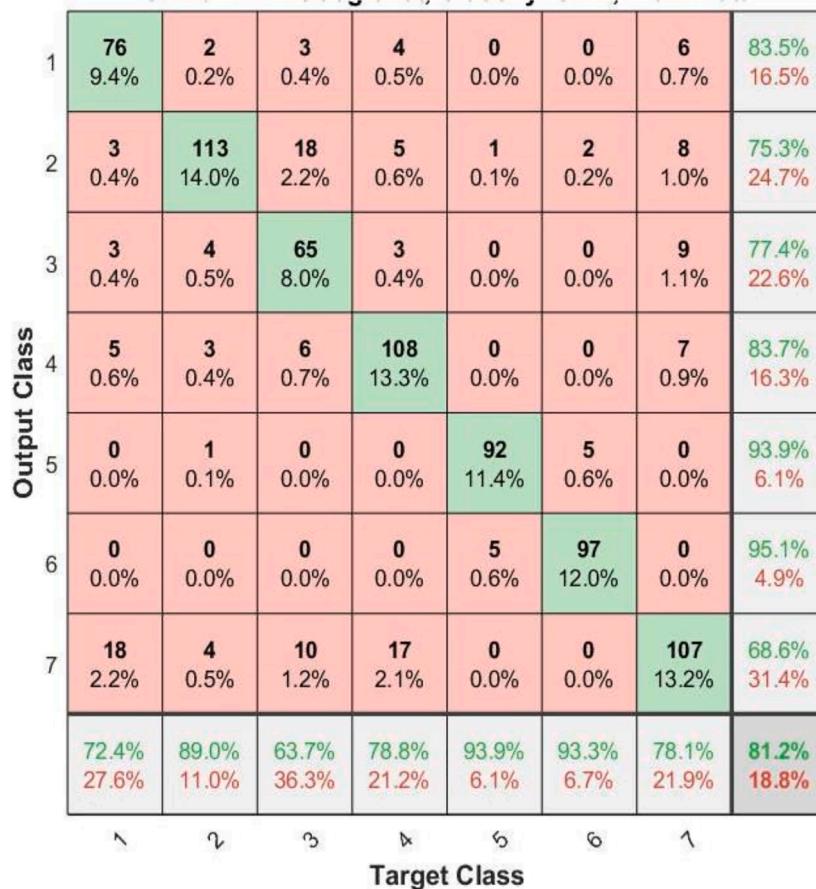


Fig. 8. Confusion matrix of the 2-class and 7-class.

proposed method the for seven class problem is demonstrated. In detail the class five (mild dysplastic), and class six (severe dysplastic), has better performance comparing other classes because of less uncertainty in the shape, size and brightness of the image cells and the other classes

has less performance.

Also, the results of two-class problem using the IFDL model are shown in Table 3. The results have been run with different pre-train networks, containing Google-Net, Alex-Net, and Res-Net50.

Table 6

Comparison of the proposed model with the counterpart methods.

Method	Accuracy %	Specificity %	Sensitivity %	AUC %
Azur machine learning (Abhinaav R, and D. Brindha, 2019)	90.18	93.56	93.33	95.43
Feature concatenation (Nguyen et al. 2018)	92.63	-	-	-
Deep Pap (Zhang, et al. 2018)	98.30	98.30	98.20	99.80
EML (IbrahimWaly et al., 2021)	97.96	-	96.99	-
Transfer learning (VGG19) with resizing (Bhatt et al. 2021)	85.15	88.22	87.24	-
Transfer learning (Res-Net) with resizing (Bhatt et al., 2021)	93.14	92.81	94.56	-
Transfer learning (Efficient-Net) with resizing (Bhatt, et al. 2021)	93.03	94.27	94.31	-
Bi-path Deep CNN (Desiani et al. 2021)	91.00	91.00	87.00	-
Deep-Cervix (Rahaman et al. 2021)	98.32	-	-	-
DCAVN (Khamparia et al. 2021)	98.1	99.2	98.3	-
Enhanced Deep Feature (Basak et al. 2021)	98.32	97.65	98.66	-
IFDL [this study]	98.51	98.80	98.45	99.95

The IFDL with Google-Net with 99.20% and 99.25% assigned the maximum accuracy and sensitivity and Res-Net with 99.99% assigned maximum of AUC. Also the sensitivity of Google-Net and Alex-Net is equal to 99.25%. It can be said that all the networks have a little difference in the competition, but overall, the Google-Net has a better performance.

Furthermore, the results of the IFDL model were compared with the basic classification methods like CNN with soft-max, Belief Network (BF) and Fuzzy Logic (FL). [Table 4](#) shows the result of the basic model comparing the IFDL model. The amount of accuracy in the IFDL model, CNN with soft-max, BF and FL were 98.51%, 90.71%, 93.39% and 89.10%, respectively. The specificity of CNN was 45.05% which assigned a minimum value. It means that the ability of the soft-max classifier to detect normal cells is lower than the other models. Also, the sensitivity was 75% which is not reliable. The proposed IFDL model with 98.80% specificity and 98.45% sensitivity seems more reliable. As shown in [Table 4](#), the model has an improvement with an average of 7%, 18%, 21% and 16% compared to the base classification method in terms of specificity, sensitivity and AUC, respectively.

The ROC curve analysis of the IFDL model is shown in [Fig. 6](#). According to [Fig. 6](#) (a), the AUC of the ROC curve for the IFDL model is 99% while for the Belief network method, [Fig. 6](#) (b), is 96%. The AUC of IFDL until 0.9 is tangential to the TP axis, while in the BF method until 0.5 is tangential to the TP axis.

The ROC curve analysis of the results of each class in seven class problem is shown in [Fig. 7](#).

In the ROC curve class five and class six have the highest area under the ROC curve. In the next level class one, two and four have the maximum area under the ROC curve more than class seven, but the class three have the minimum area under the ROC curve. As it is mentioned before the class three has the high degree of uncertainty in shape and brightness. The area under the ROC curve of class one to seven is 95.74%, 94.76%, 92.19%, 96.34%, 99.71%, 99.23%, 93.15%, respectively. Class one to seven is described in [Table 5](#).

More over the confusion matrix of the 2-class and 7-class is shown in [Fig. 8](#).

In the confusion matrix, accuracy of each class has been shown. The overall accuracy in the end of main matrix diagonal is 81.2% with 18.8%

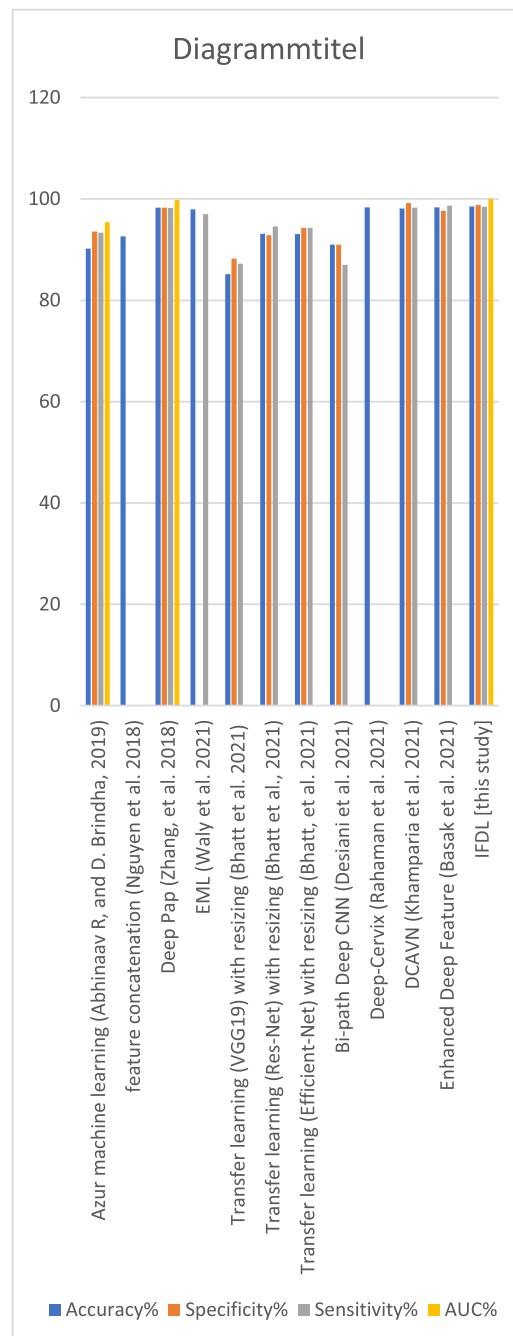


Fig. 9. Evaluation chart of the ACC, SPC, SEN and, AUC.

Table 7

Results of SIPaKMeD compare with state-of-the-art method.

Method	Accuracy %	Specificity %	Sensitivity %	AUC %
Five classifiers (Win et al. 2020)	94.09	-	-	-
Enhanced Deep (Basak et al. 2021)	97.87	98.89	98.55	-
DCNN and SVM (Plissiti et al. 2018)	93.35	98.30	98.20	99.80
CNN (Plissiti et al. 2018)	95.35	-	-	-
IFDL [This study]	97.20	98.79	98.70	99.89

error which is promising for seven class problems while in two class problem is 98.8%. Next subsection is about comparison with the other state of the art methods.

5.3. Comparison analysis with the results

The experimental results of proposed method were analyzed in previous section. To have more accurate performance assessment of the proposed algorithm, the results, first compare with pre-trained networks, pre-trained network with BF classifier. Also, the results comparison with the state of the art methods, (Abhinaav and Brindha, 2019; Nguyen et al., 2018), and (Zhang et al., 2018) is presented. Performance of the proposed method in two class problem, with the other state of the art method demonstrate in Table 6. The IFDL model in all of the metrics outperforms the Azur machine learning (Abhinaav and Brindha, 2019), feature concatenation (Nguyen et al., 2018), Deep Pap (Zhang et al., 2018) and extreme learning machine EML ((Waly, Sikkandar, Aboamer, Kadry, & Thinnukool, 2021) Waly et al., 2021). Also, the results have been compared with (Zhang, et al., 2018) in seven class problem. Also The IFDL model also outperforms the Azur machine learning model (Bhatt et al., 2021), Bi-path architecture (Anita et al., 2021), Deep-cervix (Rahaman et al., 2021), DCAVN (Khamparia et al., 2021), Enhanced Deep feature (Basak et al., 2021) studies with an average classification rate of 94.19% while the proposed method has the average accuracy of 98.51%. Also, the IFDL model outperforms its counterpart methods in terms of specificity which refers to ability of the proposed model to recognize normal cells. The IFDL model has 4.4% more specificity than the other models, with an average specificity of 94.37%. The IFDL model detects abnormal cells with 4.1% sensitivity more than the average of related works. The sensitivity of the proposed model is 98.45 which competes with related works. The IFDL model overall performance is 20% greater than the Deep Pap results reported in Zhang et al. (2018) with 98.30%, 98.30%, 98.20%, 99.80%, accuracy, specificity, sensitivity and AUC, respectively, as shown in Fig. 9.

The transfer learning model (VGG19) has the lowest efficiency. The Deep Cervix method competes with the Deep pap and the IFDL model. Overall, the transfer learning models had the lowest performance while the enhanced models showed the better performance.

The proposed IFDL method was further investigated on another public well-known dataset called (SIPaKMeD) and the results was reported in Table 7. The SIPaKMeD database consists of 4049 images of isolated cells of Pap smear slides collected by (Plissiti et al., 2018) which is used in many recent research. In this dataset normal cells consist of superficial and Parabasal cells, abnormal cells consist of Koilocytopic cells and Dyskeratotic cells, and benign cells consist of Metaplastic cells. Therefore, the classification problem is three class problem.

As shown in Table 7, the results of the IFDL model on the SIPaKMeD is competitive with the state-of-the-art models. It also outperforms related models with 98.70% and 99.89% sensitivity and AUC, respectively. The IFDL model shows 97.20% accuracy, which is more than the average accuracy of other models including Five Classifiers model (Win et al., 2020), Enhanced Deep (Basak et al., 2021), DCNN and SVM (Plissiti et al., 2018) CNN (Plissiti et al., 2018). The sensitivity of the IFDL with 98.70% further reveals ability of model to detect cancerous cells. Furthermore, the IFDL represents better performance with 98.79% and 99.89% in terms of specificity and AUC, respectively.

6. Conclusion

The key challenges in Pap smear image data analysis are uncertainty handling in correct classification and detection columnar cells and avoiding sample rejection of Belief network due to low probability of each class. The Dempster Combinational rule for uncertainty handling, and fuzzy weighting to prevent data rejection, was implemented. The dataset which is used in this study is a standard dataset, collected in Herlev University Hospital and SIPaKMeD dataset. The Herlev dataset

contains 917 and SIPaKMeD dataset consist of 4096 Pap smear cell images that is labeled by technicians and doctors. The images after data augmentation enters to pre trained networks for feature extraction then the features feed into the proposed Belief network with fuzzy weighting system. The results were evaluated by average of 5- fold cross validation and with the metrics of accuracy, specificity, sensitivity, and AUC. The results of proposed model, are compatible with the other state of the art methods. Pre-trained CNN methods (Google net, res net and Alex net) and BF method were applied to Pap images for classification of the images and compared with the proposed model. The results show 99.20%, 99.75%, 99.25%, in accuracy, specificity, and sensitivity respectively in two class problem, and in seven class the results are promising. The area under the ROC curve of 99.01% shows the superiority of our proposed method which ameliorate the other method up to an average of 10% in each metric. Also, the proposed model outperforms all of the state-of-the-art methods in Accuracy, Specificity, Sensitivity and AUC.

Credit authorship contribution statement

Mona Benhari: designing the architecture, implementing, writing, Rahil Hossini: Main Idea, supervision, writing.

Dataset available at

The Herlev dataset is available at <http://mdeLab.aegean.gr/downloads>.

The SIPaKMeD is available at https://www.cs.uoi.gr/~marina/sipa_kmed.html.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Singh, Sanjay K., & Goyal, Anjali. (2021). A stack autoencoders based deep neural network approach for cervical cell classification in pap-smear images. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 14(1). pp. 62-70(9).
- Abhinaav, R., & Brindha, Dr. D. (2019). Abnormality detection and severity classification of cells based on features extracted from papanicolaou smear images using machine learning. In *International Conference on Computer Communication and Informatics*.
- Khamparia, Aditya, Gupta, Deepak, Rodrigues, Joel J. P. C., & Albuquerque, Victor Hugo C.de (2021). DCAVN: Cervical cancer prediction and classification using deep convolutional and variational autoencoder network. *Multimedia Tools and Applications*, 80, 30399–30415.
- Bhatt, Anant R., Ganatra, Amit, & Kotecha, Ketan (2021). Cervical cancer detection in pap smear whole slide images using convNet with transfer learning and progressive resizing. *Peer j computer science*.
- Desiani, Anita, Member, IAENG, Erwin, Member, IAENG, Suprihatin, Bambang, & Yahdin, Sugandi (2021). Bi-path architecture of CNN segmentation and classification method for cervical cancer disorders based on pap-smear images. *IAENG International Journal of Computer Scienc*, 48(3).
- Tripathi, Anurag, Arora, Aditya, & Bhan, Anupama (2021). Classification of cervical cancer using Deep Learning Algorithm. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*.
- Harangi, Balazs, Toth, Janos, Bogacsovics, Gergo, Kupas, David, Kovacs, Laszlo, & Hajdu, Andras (2019). Cell detection on digitized Pap smear images using ensemble of conventional image processing and deep learning techniques. In *11th IEEE International Symposium on Image and Signal Processing and Analysis (ISPA)*.
- Taha, Bilal, Dias, Jorge, & Werghi, Naoufel (2017). *Classification of Cervical-Cancer Using Pap-Smear Images: A Convolutional Neural Network Approach*, 723 pp. 261–272). Springer International Publishing AG.
- Dharani, C., Kaviya, S., Maheshwari, S., Monisha, K., & Elayaraja, P. (2020). Visualization of cervical cancer classification using Deep Convolutional Neural

- Network. *International Journal of New Innovations in Engineering and Technology*, 2319-6319, 196–206.
- Diniz, Débora N., Rezende, Mariana T., & Bianchi, Andrea G. C. (2021). A deep learning ensemble method to assist cytopathologists in pap test image classification. *Journal of imaging*, 7(7).
- Denoeux, T. (2000). A Neural Network Classifier based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 3(2), 131–150.
- Sahba, Farhang, & Tizhoosh, Hamid R. (2003). Filter fusion for image enhancement using reinforcement learning. *Ccec 2003 - ccgei 2003 (p. 4)*. Montreal, May/mai. IEEE.
- Araujoa, Flavio H. D., Silva, Romuere R. V., Ushizima, Daniela M., Resende, Mariana T., Carneiro, Claudia M., Bianchi, Andrea G. Campos, & Medeiros, Fatima N. S. (2018). Deep learning for cell image segmentation and ranking. *Computerized Medical Imaging and Graphics*, 29.
- GS. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.
- Hansang Lee Junmo Kim. (2015). Segmentation of overlapping cervical cells in microscopic images with superpixel partitioning and cell-wise contour refinement. ISBN.
- Basak, Hritam, Kundu, Rohit, Chakraborty, Sukanta, & Das, Nibaran (2021). Cervical cytology classification using PCA and GWO enhanced deep features selection. *SN Computer Science*, 369(2).
- Martinez, Jose, Bueno-Crespo, Andre, Espana, Raquel Martinez, Remezal, Manuel, Ortiz, Ana, Ortiz, Sebastian, & Pedro, Juan (2020). Classifying Papanicolaou cervical smear through a cell Merger approach by deep learning approach. *Expert System with Application*, 31.
- Bora, Kangkana, Chowdhury, Manish, & Mahanta, Lipi B. (2016). *Pap Smear Image Classification Using Convolutional neural network*. Guwahati, India: ICVGIP, December 18-22, 2016.
- Win, Kyi Pyar, Kitajture, Yuttana, Hamamoto, Kazuhiko, & Aung, Thet Myo (2020). Computer-assisted screening for cervical cancer using digital image processing of pap smear images. *Applied Science*, 10(1800).
- Zhang, Ling, Lu, Le, Member, Senior, IEEE, Isabella Nogues, Summers, Ronald M., & Liu, Shaoxiong (2018). DeepPap: Deep convolutional networks for cervical cell classification. *Journal of Biomedical and Health Informatics*, 11.
- Nguyen, Long D., Lin, Dongyun, Lin, Zhiping, & Cao, Jiuwen (2018). Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. In *IEEE International Symposium on Circuits and Systems (ISCAS)*.
- Marcin Woźniak, Jakub Silka & Michał Wieczorek. (2021). Deep neural network correlation learning mechanism for CT brain tumor detection. Data intelligent in IoT.
- Plissiti, Marina E., Dimitrakopoulos, P., Sfikas, G., Nikou, Christophoros, Krikoni, O., & Charchanti, A. (2018). SIPAKMED: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. In *IEEE International Conference on Image Processing (ICIP)*.
- Masoud S. Nosrati and Ghassan Hamarneh. (2105). Segmentation of overlapping cervical cells: a variational method with star-shape prior. ISBI challenge. sydni.
- MamunurRahaman, Md, & ChenLi, YudongYao. (2021). DeepCervi A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. *Computers in Biology and Medicine*, 136.
- Wu, Miao WuMiao, Yan, Chuango, Liu, Huiqiang, Liu, Qian, & Yin, Yi (2019). Automatic classification of cervical cancer from cytological images by using convolutional neural network. *Bioscience Reports*, 14.
- IbrahimWaly, Mohamed, Sikkandar, Mohamed Yacin, Aboamer, Mohamed Abdelkader, Kadry, Seifedine, & Thinnukool, Orawit (2021). Optimal deep convolution neural network for cervical cancer diagnosis model. *Computers, Materials & Continua*, 70(2), 3296–3310.
- Al-batah, Mohammad Subhi, Isa, Nor Ashidi Mat, Klaib, Mohammad Fadel, & Al-Betar, and Mohammed Azmi (2014). Multiple adaptive neuro-fuzzy inference system with automatic features extraction algorithm for cervical cancer recognition. *Computational and Mathematical Methods in Medicine*, 12, Article ID 181245.
- Mohammed, Mohammed Aliy, Abdurahman, Fetulhak, & Ayalew, Yodit Abebe (2021). Single-cell conventional pap smear image classification using pre-trained deep neural network architectures. *BMC Biomedical Engineering*, 11.
- Sompawong, N., Mopan, J., Pooprasert, P., Himakhun, W., Suwanarruk, K., Ngamvirojcharoen, J., Vachiramon, T., & Tantibundhit, C. (2019). Automated Pap Smear Cervical Cancer Screening Using Deep Learning. In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*.
- Yaman, Orhan, & Tuncer, Turker (2022). Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images. *Biomedical Signal Processing and Control*, 73.
- Hosseini, Rahil, Dehmehki, Jamshid, Barman, Sarah, Mazinani, Mahdi, & Qanadli, Salah (2010). Modeling uncertainty in classification design of a computer aided. *Medical Imaging*, 7624, 8.
- Dash, Sonali, Verma, Sahil, Kavita, Savitri Bevinakoppa, Wozniak, Marcin, Shafi, Jana, & Ijaz, Muhammad Fazal (2022). Guidance image-based enhanced matched filter with modified thresholding for blood vessel extraction. *MDPI*, 14.
- Twinkle Dalal and Manjeet Singh. (2021). Review Paper on Leaf Diseases Detection and Classification Using Various CNN Techniques. Springer Nature, 140, Mobile Radio Communications and 5G Networks.
- Kumararaja, V., & Deepa, K. (2021). Pap smear image classification to predict urinary cancer using artificial neural networks. *Annals of the Romanian Society for Cell Biology*, 5(2), 1092–1098.
- Palanisamy, Vijayanand Sellamuthu, Athiappan, Rajiv Kannan, & Nagalingam, Thirugnanasambandan (2021). Pap smear based cervical cancer detection using residual neural networks deep learning architecture. *Concurrency and Computation*.
- William, Wasswa, Ware, Andrew, Basaza-Ejiri, Annabella Habinka, & Obungoloch, Johnes (2018). A review of Image Analysis and Machine Learning Techniques for Automated Cervical Cancer Screening from pap-smear images. *Computer Methods and Programs in Biomedicine*, 13.
- William, Wasswa, Ware, Andrew, Habinka, Annabella, Ejiri, Basaza, & Obungoloch, Johnes (2019). A pap smear analysis tool (PAT) for detection of cervical cancer from pap smear images. *BioMedical Engineering online*, 22.
- Huang, Wenbo, ZhangB, Lei, Wu, Hao, Min, Fuhong, Song, Aiguo, Member, Senior, & IEEE. (2021). Channel-Equalization-HAR: A light-weight convolutional neural network for wearable sensor based human activity recognition. *Transactions on Mobile Computing*.
- Cheng, Xin, Zhang, Lei, Tang, Yin, Liu, Yue, Wu, Hao, & He, Jun (2022). Real-time human activity recognition using conditionally parametrized convolutions on mobile and wearable devices. *IEEE Sensors Journal*, 22(6).
- Tang, Yin, Zhang, Lei, Min, Fuhong, & He, Jun (2022). Multi-scale deep feature learning for human activity recognition using wearable sensors. *IEEE Transactions on Industrial Electronics*.
- Tong, Zheng, Xu, Philippe, & Denmux, Thierry (2019). *Conv Net and Dempster-Shafer Theory for Object Recognition* (pp. 368–381). Springer Nature Switzerland.
- Lu, Zhi, Carneiro, Gustavo, Bradley, Andrew P., Ushizima, Daniela, Nosrati, Masoud S., Bianchi, Andrea G. C., Carneiro, Claudia M., & Hamarneh, Ghassan (2015). Evaluation of three algorithms for the segmentation of overlapping cervical cells. *Journal of Biomedical and Health Informatics*, (september 2015), 11.