

Classification of Pancreatic Cysts in Computed Tomography Images Using a Random Forest and Convolutional Neural Network Ensemble

Konstantin Dmitriev¹, Arie E. Kaufman¹, Ammar A. Javed², Ralph H. Hruban³, Elliot K. Fishman⁴, Anne Marie Lennon^{2,5}, and Joel H. Saltz⁶

¹ Department of Computer Science, Stony Brook University, Stony Brook

² Department of Surgery, Johns Hopkins School of Medicine, Baltimore, MD

³ The Sol Goldman Pancreatic Cancer Research Center, The Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD

⁴ Department of Radiology, Johns Hopkins School of Medicine, Baltimore, MD

⁵ Division of Gastroenterology and Hepatology, Johns Hopkins School of Medicine, Baltimore, MD

⁶ Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY

Abstract. There are many different types of pancreatic cysts. These range from completely benign to malignant, and identifying the exact cyst type can be challenging in clinical practice. This work describes an automatic classification algorithm that classifies the four most common types of pancreatic cysts using computed tomography images. The proposed approach utilizes the general demographic information about a patient as well as the imaging appearance of the cyst. It is based on a Bayesian combination of the random forest classifier, which learns subclass-specific demographic, intensity, and shape features, and a new convolutional neural network that relies on the fine texture information. Quantitative assessment of the proposed method was performed using a 10-fold cross validation on 134 patients and reported a classification accuracy of 83.6%.

1 Introduction

Pancreatic cancer, or pancreatic ductal adenocarcinoma (PDAC) as it is formally known, is one of the most lethal of all cancers with an extremely poor prognosis and an overall five-year survival rate of less than 9%. There are no specific early symptoms of this disease, and most of the cases are diagnosed at an advanced stage after the cancer has spread beyond the pancreas. Early detection of the precursors of PDAC could offer the opportunity to prevent the development of invasive PDAC. Two of the three precursors of PDAC, intra-ductal papillary mucinous neoplasms (IPMNs) and mucinous cystic neoplasms (MCNs), form pancreatic cysts. These cysts are common and easy to detect with currently available imaging modalities such as computed tomography (CT) and magnetic resonance imaging. IPMNs and MCNs can be relatively easily identified and offer the potential for the early identification of PDAC. However, the issue

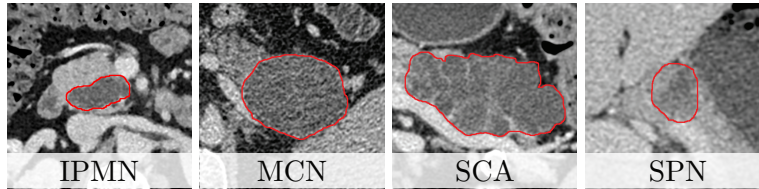


Fig. 1: Examples of pancreatic cyst appearance in CT images.

is complicated because there are many other types of pancreatic cysts. These range from entirely benign, or non-cancerous cysts, such as serous cystadenomas (SCAs), which do not require surgical intervention, to solid-pseudopapillary neoplasms (SPNs), which are malignant and should undergo surgical resection. These issues highlight the importance of correctly identifying the type of cyst to ensure appropriate management [5].

The majority of pancreatic cysts are discovered incidentally on computed tomography (CT) scans, which makes CT the first available source of imaging data for diagnosis. A combination of CT imaging findings in addition to general demographic characteristics, such as patient age and gender, are used to discriminate different types of pancreatic cysts [5]. However, correctly identifying cyst type by manual examination of the radiological images can be challenging, even for an experienced radiologist. A recent study [9] reported an accuracy of 67-70% for the discrimination of 130 pancreatic cysts on CT scans performed by two readers with more than ten years of experience in abdominal imaging.

The use of a computer-aided diagnosis (CAD) algorithm may not only assist the radiologist but also ameliorate the reliability and objectivity of differentiation of various pancreatic cysts identified in CT scans. Although many algorithms have been proposed for the non-invasive analysis of benign and malignant masses in various organs, to our knowledge, there are no CAD algorithms for classifying pancreatic cyst type. This paper presents a novel non-invasive CAD method for discriminating pancreatic cysts by analyzing imaging features in conjunction with patient’s demographic information.

2 Data Acquisition

The dataset in this study contains 134 abdominal contrast-enhanced CT scans collected with a Siemens SOMATOM scanner (Siemens Medical Solutions, Malvern, PA). The dataset consists of the four most common pancreatic cysts: 74 cases of IPMNs, 14 cases of MCNs, 29 cases of SCAs, and 17 cases of SPNs. All CT images have 0.75mm slice thickness. The ages of the subjects (43 males, 91 females) range from 19 to 89 years (mean age 59.9 ± 17.4 years).

One of the most critical parts in the computer-aided cyst analysis is segmentation. The effectiveness and the robustness of the ensuing classification algorithm depend on the precision of the segmentation outlines. The outlines of each cyst (if multiple) within the pancreas were obtained by a semi-automated

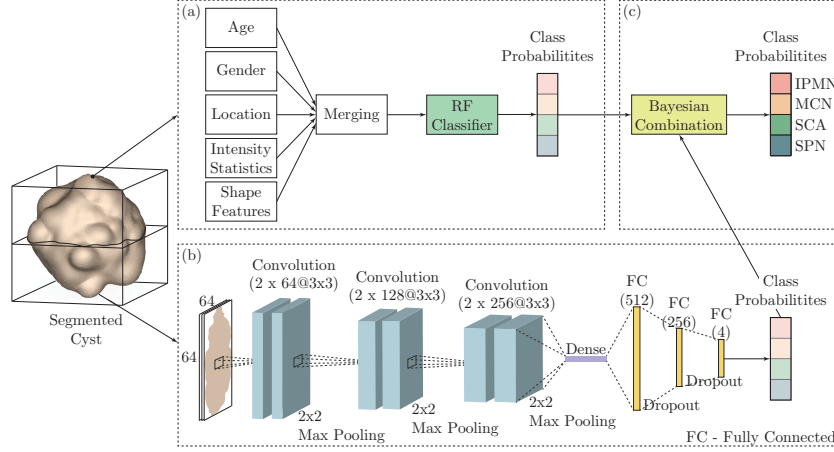


Fig. 2: A schematic view of the proposed classification ensemble of (a) a random forest trained to classify vectors of quantitative features, and (b) a convolutional neural network for classification based on the high-level imaging features. Their Bayesian combination (c) generates the final class probabilities.

graph-based segmentation technique [3] (Fig. 1), and were confirmed by an experienced radiologist (E.F.). The histopathological diagnosis for each subject was confirmed by a pancreatic pathologist (R.H.H.) based on the subsequently resected specimen. The segmentation step was followed by a denoising procedure using the state-of-the-art BM4D enhancement filter [6].

3 Method

This work describes an ensemble model, designed to provide an accurate histopathological differentiation for pancreatic cysts. This model consists of two principal components: (1) a probabilistic random forest (RF) classifier, which analyzes manually selected quantitative features, and (2) a convolutional neural network (CNN) trained to discover high-level imaging features for a better differentiation. We propose to analyze 2D axial slices, which can be more efficient in terms of memory consumption and computation compared to the analysis of 3D volumes. The overall schema of the proposed method is illustrated in Fig. 2.

3.1 Quantitative Features and Random Forest

The most common features mentioned in the medical literature that are used for initial pancreatic cyst differentiation involve gender and age of the subject, as well as location, shape and general appearance of the cyst [9]. In this paper, we define a set \mathcal{Q} of 14 quantitative features to describe particular cases by:

(1) age $a \in \mathcal{Q}$ and gender $g \in \mathcal{Q}$ of the patient, (2) cyst location $l \in \mathcal{Q}$, (3) intensity $\mathcal{I} \subset \mathcal{Q}$ and (4) shape $\mathcal{S} \subset \mathcal{Q}$ features of a cyst. The importance and discriminative power of these features are described below.

1. **Age and Gender.** Several studies reported a strong correlation between age and gender of a patient and certain types of pancreatic cysts [5, 1]. For example, MCN and SPN often present in women of premenopausal age. In contrast, IPMNs have an equal distribution between men and women, and typically present in patients in their late 60s.
2. **Cyst location.** Certain cyst types are found in particular locations within the pancreas. For example, the vast majority of MCNs arise in the body or tail of the pancreas.
3. **Intensity features.** Due to the differences in the fine structure of pancreatic cysts, such as homogeneity versus common presence of septation, calcification or solid component, we use the set $\{\bar{I}, s, \kappa, \gamma, M\} \in \mathcal{I}$, which are the mean, standard deviation, kurtosis, skewness and median of intensities, respectively, as the global intensity features for coarse initial differentiation.
4. **Shape features.** Pancreatic cysts also demonstrate differences in shape depending on the category. Specifically, cysts can be grouped into three categories: smoothly shaped, lobulated and pleomorphic cysts [1]. To capture different characteristics of the shape of a cyst, we use volume $V \in \mathcal{S}$, surface area $SA \in \mathcal{S}$, surface area-to-volume ratio $SA/V \in \mathcal{S}$, rectangularity $r \in \mathcal{S}$, convexity $c \in \mathcal{S}$ and eccentricity $e \in \mathcal{S}$ features summarized in [11].

Given a set $\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_k, y_k)\}$ of examples \mathbf{x}_i of pancreatic cysts of known histopathological subtypes $y_i \in \mathcal{Y} = \{IPMN, MCN, SCA, SPN\}$, we compute a concatenation $\mathbf{q}_i = (a_i, g_i, l_i, \bar{I}_i, s_i, \kappa_i, \gamma_i, M_i, V_i, S_i, SA_i, SA/V_i, r_i, c_i, e_i)$ of the described features for all k samples in the set \mathcal{D} .

Following feature extraction, we use an RF classifier to perform the classification of a feature vector \mathbf{q}_m computed for an unseen cyst sample \mathbf{x}_m . RF-based classifiers have shown excellent performance in various classification tasks, including numerous medical applications, having high accuracy of prediction and computation efficiency [7, 8].

More formally, we use a forest of T decision trees implemented with the scikit-library¹. Each decision tree θ_t predicts the conditional probability $P_{\theta_t}(y|\mathbf{q}_m)$ of histopathological class y , given a feature vector \mathbf{q}_m . The final RF class probability can be found as the following:

$$\tilde{P}_1(y_m = y|\mathbf{x}_m) = \tilde{P}_{\text{RF}}(y_m = y|\mathbf{q}_m) = \frac{1}{T} \sum_{t=1}^T P_{\theta_t}(y_m = y|\mathbf{q}_m). \quad (1)$$

For more details, we refer the reader to the technical report [2].

3.2 CNN

As described in Sec. 4, RF trained on the proposed quantitative features can be used for cyst classification with reasonably high accuracy. However, despite

¹ <http://scikit-image.org>

high generalization potential, the proposed features do not take full advantage of the image information. In particular, due to variations in the internal structure of pancreatic cysts, they show different image characteristics: SCA often has a honeycomb-like appearance with a central scar or septation, MCN demonstrates a “cyst within cyst” appearance with peripheral calcification, IPMN tends to have a “cluster of grapes” appearance, and SPN typically consists of solid and cystic components [12]. However, these imaging features can overlap, especially when the cyst is small and the internal architecture cannot be differentiated.

We apply CNN as a second classifier, which can better learn barely perceptible yet important image features [10]. The proposed CNN, shown in Fig. 2(b), contains 6 Convolutional, 3 Max-pooling, 2 Dropout and 3 Fully-connected (FC) layers. Each convolutional and the first two FC layers are followed by the rectified linear unit (ReLU) activation function; the last FC layer ends with the softmax activation function to obtain the final class probabilities.

The data for training and testing the proposed CNN were generated as follows. Each 2D axial slice X_{ij}^{Slice} of the original 3D bounding box $\{X_{ij}^{\text{Slice}}\}$ with a segmented cyst \mathbf{x}_i was down-/up-sampled to 64×64 pixels squares, using bicubic interpolation. Visual examination confirmed the preservation of the important features. Due to the generally spherical shape of a cyst, slices near the top and the bottom of the volume do not contain enough pixels of a cyst to make an accurate diagnosis. Therefore, slices with the overlap ratio less than 40%, defined as the percentage of cyst pixels in a slice, were excluded. We also incorporated a data augmentation routine to increase the size of the training dataset and to prevent over-fitting: (1) random rotations within $[-25^\circ; +25^\circ]$ degree range; (2) random vertical and horizontal flips; (3) and random horizontal and vertical translations within $[-2; +2]$ pixels range.

The network was implemented using the Keras library² and trained on 512-sized mini-batches to minimize the class-balanced cross-entropy loss function using Stochastic Gradient Descent with a 0.001 learning rate, momentum of 0.9, weight decay of 0.0005 for 100 epoch. In the testing phase, each slice with the overlap ratio more than 40% was analyzed by the CNN separately, and the final probabilities were obtained by averaging the class probabilities for each slice:

$$\tilde{P}_2(y_m = y | \mathbf{x}_m) = \tilde{P}_{\text{CNN}}(y_m = y | \{X_{ij}^{\text{Slice}}\}) = \frac{1}{J_m} \sum_{j=1}^{J_m} P_{\text{CNN}}(y_m = y | X_{mj}^{\text{Slice}}), \quad (2)$$

where $P_{\text{CNN}}(y_m = y | X_{mj}^{\text{Slice}})$ is the vector of class probabilities, and J_m is the number of 2D axial slices used for the classification of cyst sample \mathbf{x}_m .

3.3 Ensemble

Although our dataset is representative of the types of cysts that arise in the population, we still recognize that it contains limited information and might not include enough cases of cysts of rare imaging appearance, which is crucial

² Chollet, F.: Keras. <https://github.com/fchollet/keras> (2015)

Table 1: Confusion matrices of the RF (left) and CNN (right) classifiers

Ground Truth	RF Prediction (%)				Ground Truth	CNN Prediction (%)			
	<i>IPMN</i>	<i>MCN</i>	<i>SCA</i>	<i>SPN</i>		<i>IPMN</i>	<i>MCN</i>	<i>SCA</i>	<i>SPN</i>
<i>IPMN</i>	95.9	1.4	2.7	0.0	<i>IPMN</i>	93.2	4.0	1.4	1.4
<i>MCN</i>	21.4	64.3	14.3	0.0	<i>MCN</i>	57.1	28.6	14.3	0.0
<i>SCA</i>	51.7	3.5	37.9	6.9	<i>SCA</i>	37.9	0.0	48.3	13.8
<i>SPN</i>	5.9	0.0	0.0	94.1	<i>SPN</i>	0.0	0.0	0.0	100.0

for obtaining robust CNN performance. Therefore, we hypothesize that the RF classifier will show a better performance at classifying small cysts, which do not have enough distinctive imaging features, by utilizing the clinical information about the patient and the general intensity and shape features, whereas CNN is expected to show a similar performance but at analyzing large cysts.

It has been shown that combinations of multiple classifiers, classifier ensembles, achieve superior performance compared to single classifier models [4], by learning different, presumably independent classification subproblems separately. Therefore, after training RF and CNN classifiers independently, we perform a Bayesian combination to ensure that a more robust and accurate classifier has more power in making the final decision. Mathematically, the final histopathological diagnosis \hat{y} can be written in the following way:

$$\hat{y}_m = \arg \max_{y \in \mathcal{Y}} \frac{\tilde{P}_1(y_m = y | \mathbf{x}_m) \tilde{P}_2(y_m = y | \mathbf{x}_m)}{\sum_{y' \in \mathcal{Y}} \prod_{c=1}^2 \tilde{P}_c(y_m = y' | \mathbf{x}_m)}. \quad (3)$$

4 Results and Discussion

We evaluated the performance of the proposed method using a stratified 10-fold cross-validation strategy, maintaining similar data distribution in training and testing datasets to avoid possible over- and under-representation of classes due to the imbalance in the dataset. Classification performance is reported in terms of the normalized averaged confusion matrix and the overall classification accuracy. We also analyze the dependency between the accuracy of the individual and ensemble classifiers and the average size of the misclassified cysts.

All experiments were performed using an NVIDIA Titan X (12GB) GPU. The training of RF and CNN classifiers took approximately 1 second and 30 minutes, respectively, during each cross-validation loop. The test time for the final class probabilities took roughly 1 second to compute for a single sample.

Results of the individual classifiers. We first compare the performance of the RF and CNN classifiers separately, and the overall accuracy is 79.8% and 77.6%, respectively. The quantitative details are provided in Table 1. The experiments showed that the accuracy of 30 trees in RF lead to the error convergence and was sufficient to achieve the best performance. Prior to developing the proposed set of quantitative features, we also evaluated the performance of the RF

Table 2: Confusion matrix of the final ensemble classifier.

Ground Truth	Ensemble Prediction (%)			
	<i>IPMN</i>	<i>MCN</i>	<i>SCA</i>	<i>SPN</i>
<i>IPMN</i>	95.9	1.4	1.4	1.4
<i>MCN</i>	14.3	64.3	21.4	0.0
<i>SCA</i>	34.5	3.5	51.7	10.3
<i>SPN</i>	0.0	0.0	0.0	100.0

classifier when using only age, gender, and the location of the cyst within the pancreas, as the most objective criteria used by clinicians. The overall accuracy was 62%, and adding the volume of the cyst as a feature improved the classification by 2.2%. In addition, we investigated the performance advantages for the CNN when using the data augmentation routine. Specifically, we found that the use of data augmentation improves the overall accuracy of the CNN by 13.2%.

One of the interesting, but also expected, outcomes is the average size of the misclassified cysts. In particular, the CNN classifier struggles to correctly interpret cysts of a volume smaller than 9 cm³ or 2.3 cm in diameter (average volume and diameter of misclassified cysts are 5.1 cm³ and 1.3 cm, respectively), which are reasonably challenging due to the absence of distinctive appearance. However, the accuracy of the RF does not show such dependence (average volume and diameter of misclassified cysts are 81 cm³ and 5.2 cm, respectively).

Results of the ensemble classifier. In this experiment, we test the effect of the Bayesian combination of the RF and CNN classifiers on the performance, and the results are presented in Table 2. The overall accuracy is 83.6%, which is higher than the performance of the individual classifiers. It is also interesting to note the change in the average volume and diameter of the misclassified cysts, which are 65 cm³ and 4.8 cm for the ensemble model, respectively. These results validate our hypothesis and justify the decision to combine the RF and CNN classifiers into a Bayesian combination to consider their separate diagnoses depending on how accurate they have been at analyzing the training dataset.

5 Conclusion and Future Work

In this work, we proposed an ensemble classification model to identify pancreatic cyst types automatically. The proposed algorithm is based on a Bayesian combination of an RF classifier and a CNN to make use of both clinical information about the patient and fine imaging information from CT scans. The reported results showed promising performance and achieved an overall accuracy of 83.6%. However, our study faces some limitations. In particular, our dataset was limited to only four most common pancreatic cyst types. Future work will extend the model to include other types and will evaluate the ability of the algorithm to differentiate IPMNs and MCNs with low- or intermediate-grade dysplasia from those with high-grade dysplasia or an associated invasive adenocarcinoma. This differentiation is critical in determining appropriate therapy.

Acknowledgments

This research has been generously supported by The Marcus Foundation, Inc., and partially by NSF grants CNS0959979, IIP1069147, CNS1302246, NRT1633299, CNS1650499, IIS1527200, and NIH grant CA62924.

References

1. Cho, H.W., Choi, J.Y., Kim, M.J., Park, M.S., Lim, J.S., Chung, Y.E., Kim, K.W.: Pancreatic tumors: emphasis on CT findings and pathologic classification. *Korean J Radiol* 12(6), 731–739 (2011)
2. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Microsoft Research Cambridge, Tech. Rep. MSRTR-2011-114 5(6), 12 (2011)
3. Dmitriev, K., Gutenko, I., Nadeem, S., Kaufman, A.: Pancreas and cyst segmentation. *Proc. of SPIE Medical Imaging* pp. 97842C–97842C (2016)
4. Ingallhalikar, M., Parker, W., Bloy, L., Roberts, T., Verma, R.: Using multiparametric data with missing features for learning patterns of pathology. *Proc. of International Conference on MICCAI* pp. 468–475 (2012)
5. Lennon, A.M., Wolfgang, C.L., Canto, M.I., Klein, A.P., Herman, J.M., Goggins, M., Fishman, E.K., Kamel, I., Weiss, M.J., Diaz, L.A., et al.: The early detection of pancreatic cancer: What will it take to diagnose and treat curable pancreatic neoplasia? *Cancer Res.* 74(13), 3381–3389 (2014)
6. Maggioni, M., Katkovnik, V., Egiazarian, K., Foi, A.: Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.* 22(1), 119–133 (2013)
7. Raman, S.P., Chen, Y., Schroeder, J.L., Huang, P., Fishman, E.K.: CT texture analysis of renal masses: pilot study using random forest classification for prediction of pathology. *Acad Radiol* 21(12), 1587–1596 (2014)
8. Raman, S.P., Schroeder, J.L., Huang, P., Chen, Y., Coquia, S.F., Kawamoto, S., Fishman, E.K.: Preliminary data using computed tomography texture analysis for the classification of hypervascular liver lesions: generation of a predictive model on the basis of quantitative spatial frequency measurements - a work in progress. *J Comput Assist Tomogr* 39(3), 383–395 (2015)
9. Sahani, D.V., Sainani, N.I., Blake, M.A., Crippa, S., Mino-Kenudson, M., del Castillo, C.F.: Prospective evaluation of reader performance on mdct in characterization of cystic pancreatic lesions and prediction of cyst biologic aggressiveness. *AJR Am J Roentgenol* 197(1), W53–W61 (2011)
10. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imag.* 35(5), 1285–1298 (2016)
11. Yang, M., Kpalma, K., Ronsin, J.: A survey of shape feature extraction techniques. *Pattern Recogn* pp. 43–90 (2008)
12. Zaheer, A., Pokharel, S.S., Wolfgang, C., Fishman, E.K., Horton, K.M.: Incidentally detected cystic lesions of the pancreas on CT: review of literature and management suggestions. *Abdom Imaging* 38(2), 331–341 (2013)