Analysis of a Feature-Deselective Neuroevolution Classifier (FD-NEAT) in a Computer-Aided Lung Nodule Detection System for CT Images

Maxine Tan
Dept. of Electronics & Informatics
Vrije Universiteit Brussel
Brussel, Belgium
Interdisciplinary Institute for Broadband Technology (IBBT),
Dept. of Future Media and Imaging (FMI), Ghent, Belgium
mytan@etro.vub.ac.be

Rudi Deklerck
Dept. of Electronics & Informatics
Vrije Universiteit Brussel
Brussel, Belgium
Interdisciplinary Institute for Broadband Technology (IBBT),
Dept. of Future Media and Imaging (FMI), Ghent, Belgium
rdeklerc@etro.vub.ac.be

Bart Jansen
Dept. of Electronics & Informatics
Vrije Universiteit Brussel
Brussel, Belgium
Interdisciplinary Institute for Broadband Technology (IBBT),
Dept. of Future Media and Imaging (FMI), Ghent, Belgium
bjansen@etro.vub.ac.be

Jan p. Cornelis
Dept. of Electronics & Informatics
Vrije Universiteit Brussel
Pleinlaan 2
B-1050 Brussel, Belgium
jpcornel@etro.vub.ac.be

ABSTRACT

Systems for Computer-Aided Detection (CAD), specifically for lung nodule detection received increasing attention in recent years. This is in tandem with the observation that patients who are diagnosed with early stage lung cancer and who undergo curative resection have a much better prognosis. In this paper, we analyze the performance of a novel feature-deselective neuroevolution method called FD-NEAT to retain relevant features derived from CT images and evolve neural networks that perform well for combined feature selection and classification. Network performance is analyzed based on radiologists' ratings of various lung nodule characteristics defined in the LIDC database. The analysis shows that the FD-NEAT classifier relates well with the radiologists' perception in almost all the defined nodule characteristics, and shows that FD-NEAT evolves networks that are less complex than the fixed-topology ANN in terms of number of connections.

Categories and Subject Descriptors: I.2.6 [**Artificial Intelligence**]: Learning – *Connectionism and neural nets*

General Terms: Experimentation

Keywords: Feature selection, Neural networks, Genetic algorithms, Lung nodule detection, Medical image analysis

GECCO'12 Companion, July 7–11, 2012, Philadelphia, PA, USA.

1. INTRODUCTION

Recently, systems for Computer-Aided Detection (CAD) have been gaining in popularity [1-3]. A significant focus on lung cancer is observed because it is by far the leading cause of cancer death among both men and women in the United States [4]. It is estimated that lung cancer will account for about 28% of all cancer deaths in the American Cancer Society's most recent estimates for lung cancer in the United States in 2009 [4]. The prognosis for the newly diagnosed with lung cancer is poor: the overall five-year survival rate is approximately 14% [5]. The primary reason for the optimism in CAD is the observation that those patients who are diagnosed with early stage lung cancer and who undergo curative resection have a much better prognosis, with five-year survival rates rising to 40 to 70% [6, 7].

Another motivation for the development of CAD systems for lung nodule detection is that screening studies, such as the Early Lung Cancer Action Project (ELCAP) in New York [8], at the Mayo clinic [9] and in Japan [10] show that CT imaging is much more sensitive in detecting nodules and lung cancers than chest radiography. The ELCAP studies also showed that the vast majority of lung cancers uncovered during screening were early stage cancers and thus, presumably had a better prognosis. The ELCAP and Japanese studies used single detector scanners with 10-mm-thick slices. In recent years, multidetector scanners use much thinner slices: 1 to 5-mm-thick slices [11], and later studies will undoubtedly utilize even thinner slices. These advancements in technology have significant implications for the detection of small and early stage cancers. They also have significant implications for radiologists. The multidetector scanners routinely generate 200 to 400 images per patient, which must be reviewed on a computer monitor. This significantly increases the reading burden on the radiologist.

Early results from the National Lung Screening Trial (NLST) conducted by the U.S. National Cancer Institute (NCI) released in 2010 [12] also show that low-dose CT screening for lung cancer cuts lung cancer death by more than 20%. The trial followed more than 53,000 current and former smokers who were randomized to have either X-ray or low-dose CT lung screening. The NLST study design is described in an article that was published in Radiology [13]. All these factors combined, explain the considerable interest in the development of CAD methods to assist radiologists in the early detection of lung cancer.

In many practical applications and especially CAD methods, the selection of relevant features is a task of tremendous importance. The urgency is to select features that are *relevant* to the system, features that should be included for the system to achieve optimal performance. The exclusion of these features will lead to sub-optimal results whereas the inclusion of *irrelevant* or *redundant* features adds unnecessary dimensions to the search space.

Many different features have been proposed for the lung nodule detection task, and various classifiers and feature selection algorithms have been analyzed. The features that have been proposed in the literature include gradient field descriptors, invariant features, shape and regional descriptors [1-3, 17-18]. In [14], we presented a novel feature deselection method called Feature Deselective NeuroEvolution of Augmenting Topologies (FD-NEAT). The method is a feature-deselective version of the NeuroEvolution of Augmenting Topologies (NEAT) algorithm proposed by Stanley and Miikkulainen [15]. We analyzed the usage and relevance of FD-NEAT in a novel computerized lung nodule detection system for CT images, and compared its performance with the more-established SVM and fixed-topology Artificial Neural Network (ANN) classifiers in [1].

In this paper, we perform an analysis of the networks evolved by FD-NEAT for the lung nodule detection task, and stratify the results according to four radiologists' ratings of the nodule characteristics in the Lung Image Database Consortium (LIDC) database [16]. A "blinded" and "unblinded" reviewing procedure was established for the database. In the blinded review stage, each radiologist individually annotated all nodules between 3 to 30 mm in diameter in a blinded fashion. Then, the annotations of the other radiologists were compiled and presented to the radiologists again for the unblinded reading, in which each radiologist re-examined the cases, but this time with the annotations of the other radiologists as additional information. No forced consensus was imposed in this unblinded review. For every CT scan, the unblinded read annotations from each of the four radiologists were combined into a single XML file.

With the data compiled by LIDC, an analysis can be performed taking into account the agreement levels between the four radiologists, for different nodule types after the unblinded read session. There are various methods of building a reference standard based on the annotations of the four radiologists. We used the method employed in [17, 18] called *ground truth with agreement level j*, the list of all the nodules marked by at least *j* of the four radiologists. In general, it is desirable to obtain higher sensitivities for nodules identified at higher agreement levels.

Following the rule-of-thumb that is usually applied for splitting a dataset in a training and independent test set – approximately ⅔ versus ⅓ of the cases – we chose to include in the training set 235 randomly-selected CT scans and in the test set 125 scans from the database [19, 20].

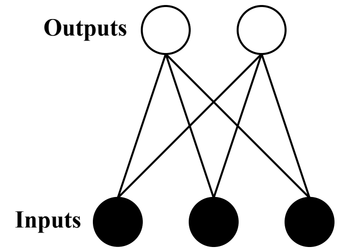

Figure 1: Initial network topology of both regular NEAT and FD-NEAT – The difference between the two methods is that with FD-NEAT, a mutation operator is implemented that enables dropping initial input connections in the algorithm.

The remainder of the paper is organized as follows: Section 2 describes the methods used. Section 3 explains the experimental framework of the CAD system. The results are presented in Section 4. In Section 5, an analysis of the networks evolved by FD-NEAT in the experiments is performed and Section 6 presents the conclusions of the study.

2. METHODS

This section provides the background of the FD-NEAT method, and the classifiers used for comparison with FD-NEAT, namely Support Vector Machines (SVM) and fixed-topology ANNs.

2.1 Feature Deselective NeuroEvolution of Augmenting Topologies (FD-NEAT)

We presented a novel feature-deselective classifier, FD-NEAT in [14]. FD-NEAT is based on NEAT, which is a method that uses genetic algorithms (GAs) to evolve the topology and the weights of networks that best fit the complexity of the task at hand. In NEAT, evolution starts from an almost minimal structure and new structure is added incrementally through the mutation operators.

Starting minimally helps NEAT to learn faster as it searches for optimal solutions over a lower-dimensional search space. NEAT only jumps to a larger search space when performance in the smaller one stagnates or does not improve over a specific number of generations. Since only additional structures that improve performance are likely to be retained, NEAT tends to discover smaller networks without superfluous structures.

NEAT and FD-NEAT both begin with a uniform population of simple networks with no hidden nodes and inputs connected directly to all the outputs (see Figure 1). It is often uncertain that all inputs are relevant and hence the initial connections of NEAT might impair the performance of the search algorithm when the search space is increased by the high number of initial input connections. The only difference between NEAT and FD-NEAT is that a mutation operator is introduced in FD-NEAT that enables irrelevant or redundant inputs in the initial population to be dropped.

FD-NEAT prunes features in the input feature set instead of adding them. Various feature selection algorithms in the literature start with a randomly-selected feature set [21], and with all features connected/ selected [22]. FD-NEAT is able to optimize its weights quickly so as to assign suitable weights to the features based on their relevance or redundancy. FD-NEAT improves on regular NEAT in that a mutation operator is introduced in FD-NEAT that enables discarding irrelevant or redundant inputs that impair network performance. In addition, at the end of evolution, less complex networks are obtained, that can be trained faster and have a lower computational complexity in the operational phase.

FD-NEAT's performance was previously examined on several simple feature selection experiments [14], e.g. the classical "exclusive or" classification problem and maneuvering a robotic car around a race track by selecting relevant sensors in a race car simulator environment (RARS) [28]. The pseudo-code of FD-NEAT is given in Table 1. The implemented FD-NEAT algorithm is slightly modified from the Matlab® NEAT algorithm by Christian Mayr [23].

2.2 Performance Comparisons of FD-NEAT with the SVM and ANN Classifiers

In our experiments, FD-NEAT is compared with two other established classifiers, namely SVMs and fixed-topology ANNs. All results are obtained for a test set that is kept completely separated from the training set.

A *LIBSVM classifier* [24] with the radial basis function (RBF) kernel, defined as $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$, $\gamma > 0$ is trained on a training set of instance-label pairs $(\mathbf{x}_i, y_i), i = 1,...,l$ where $\mathbf{x}_i \in R^n$ and $\mathbf{y} \in \{1,-1\}^l$. We used five-fold cross-validation with parallel "grid-search" [25] to determine the optimal SVM parameters. In our approach, we performed linear normalization of all the input features.

The second classifier is the *standard feed-forward ANN* having a single hidden layer with the hyperbolic tangent activation function at the hidden nodes, and the linear activation function at the single output node. The number of input nodes is equal to the number of input features (45 altogether). Only one output node is used. Values at the output node above a certain threshold value, which is chosen to obtain an acceptable false positive (FP) rate, correspond to nodule detections, and values below to non-nodule detections. The *fixed-topology ANN* is trained with the Levenberg-Marquardt backpropagation algorithm [26, 27]. The network's performance is analyzed for 5 to 40 hidden nodes. The ANN is always initialized with random weights.

For *FD-NEAT*, the hyperbolic tangent activation function is used at the hidden nodes whereas the modified sigmoidal activation function is used at the output node. Experiments show that FD-NEAT performs better when the modified sigmoidal activation function instead of the linear activation function is used at the output node.

FD-NEAT and the fixed-topology ANN (The fixed-topology ANN performs best with 11 neurons in the hidden layer) are trained on a target vector with different values (0.55, 0.7, 0.85, and 1) assigned to nodules at different agreement levels (j=1,…4). SVM is trained on a target vector in which the same target value is used for all agreement levels. This implementation was based on initial experimental results showing optimal performance under these learning strategies.

FD-NEAT is found to perform best when variance normalization (normalization of the mean and standard deviation of the training set) is applied to the input features. SVM and fixed-topology ANN perform better when the features are normalized linearly to the range [0, 1] (refer to the results of Table 2).

Table 1. Pseudo-code of FD-NEAT

```
Set parameters
Create initial population
Put first individual in species one, update species record
for each individual from the population
    while individual is not assigned to an existing species
    or there are no more species to test it against
        Compute compatibility distance, δ to compare
        with the species compatibility threshold, δ_t.
        if δ < δ_t
            Assign individual to existing species
        else
            Create a new species, update the species
            record and use individual as the reference
            for the new species
        end if
    end while
end for
generation = 1            //1st generation has passed
while generation number < maximum generations allowed
AND fitness < maximum fitness
    Evaluate the fitness of each individual
    for each species in the species record
        Compute the average and maximum fitness of
        each species and store it in the species generation
        record
    end for
    Check "refocus" parameter to see if the maximum
    overall fitness of the population has changed
    if generation number < maximum generations
    allowed AND fitness < maximum fitness
        Call reproduction function with parameters
        (Mutation operator in the reproduction function
        required for FD-NEAT function that enables
        the initial input connections to be dropped. At
        the end of the reproduction function, the new
        population, species record and innovation
        record are updated and returned to the main
        function)
    end if
    generation ++         // Increment generational counter
end while
```

3. EXPERIMENTS (FRAMEWORK OF THE LUNG CAD SYSTEM)

We presented a novel CAD system for lung nodule detection in CT images in [1]. In this section, the general framework of the CAD system is given; for a detailed description, the reader is referred to [1]. The CAD system performs several main tasks, namely preprocessing, nodule candidate detection, feature selection, and classification.

3.1 Preprocessing

In the preprocessing stage, isotropic resampling of the data to a voxel dimension of 1 mm^3 was performed. Trilinear interpolation was used to compute the grey-values between the voxel locations.

A simple lung segmentation procedure based on thresholding and a morphological closing operation similar to that proposed in [17] is implemented to segment the lung regions from the background. The lung segmentation procedure is performed to confine all subsequent computations to the lung regions only.

3.2 Nodule Candidate Detection

Specific segmentation algorithms with different parameter settings are derived for the three different nodule types, namely isolated, juxtavascular (or vessel-connected) and juxtapleural (or pleura-connected) nodules. The nodule candidate detection procedure consists of three stages: (1) Seed point detection by computation of the divergence of the normalized gradient; (2) Multiscale nodule and vessel enhancement filtering and (3) Cluster merging.

In the first stage of the nodule candidate detection procedure, we compute the divergence of the normalized gradient of the image in 3D, $k = div(\vec{w})$ where $\vec{w} = \frac{\vec{\nabla} L}{\|\vec{\nabla} L\|}$ and L is the image intensity. We do this to estimate the location of the nodule seed point.

Computing the divergence is just the first step of our proposed nodule segmentation procedure. To segment the nodule candidates, we use nodule and vessel enhancement filters proposed by Li et al. [3]. The procedure and parameters of the nodule segmentation algorithm differ slightly for the three different nodule types. Details can be found in [1].

The third and final stage of the nodule candidate detection consists of merging overlapping clusters to ensure that a single nodule is represented by a single detection rather than by two or more detections. Many nodule candidates are generated at the detection stage that will be filtered by a process of feature selection and classification.

3.3 Feature Selection and Classification

We propose features that are invariant under the group of orthogonal transformations (translations, rotations), namely features that are calculated in a 3D gauge coordinates system [29]. Any derivative expressed in the gauge coordinates is an orthogonal invariant [30]. In our experiments, we found that the gauge derivatives in the principle directions, L_{uu} and L_{vv} were good at nodule and blood vessel differentiation.

Apart from the gauge coordinate invariant features, other shape (compactness features, elongation factor, bounding ellipsoid) and regional (mean, median, standard deviation, etc.) descriptors are computed on the nodule candidate segmentations and on spherical kernels centered at the nodule candidate centroids of different radii and at different scales. Altogether, 45 features are given to the classification stage of the CAD system.

The number of detections is very large on the training dataset: 111,906 detections altogether. The nodules that are identified by different radiologists are divided into four subsets, with the number of nodules equal to 202/ 113/ 104/ 155 nodules annotated by 1/ 2/ 3/ 4 radiologists, respectively.

There are altogether 111,332 non-nodule regions in the training set generated by the detection stage of the CAD system, i.e. an average of 474 non-nodules per scan. The training and validation procedure does not work well, if there are too many negative examples (i.e. non-nodule regions) compared to positive ones. A way of reducing the number of non-nodule regions without altering their distribution in the feature space is by using a 2D self-organizing map (SOM) [31, 32].

As in any GA procedure, several runs or repetitions have to be conducted to find the optimal FD-NEAT network over the entire search space. The networks are evolved in a population of 200 networks for 200 generations. The best network, namely the network with the highest fitness (computed using a fitness function based on the classification accuracy defined in [15] and used in [14]) on the training set at the end of the evolutionary process over ten runs is used to classify the nodules in the test set. Additional details about specific parameters that are used in the FD-NEAT algorithm are given in the Appendix and in [1].

4. RESULTS

The CAD system sensitivities at all agreement levels for the FD-NEAT classifier, fixed-topology ANN and SVM classifiers, respectively, obtained at a FP rate of 4 per scan, which is an acceptable FP rate used by other CAD systems in the literature [17, 18] are given in Table 2. It can be observed that FD-NEAT's performance is highest at agreement level 3. The fixed-topology ANN performs best on nodules with agreement level 1. FD-NEAT and the fixed-topology ANN both achieve the highest sensitivity of 87.5% on nodules with agreement level 4.

Table 2. CAD system sensitivities of FD-NEAT, the fixed-topology ANN and the SVM classifiers at a FP rate of 4/scan for nodules at the four agreement levels

Agreement level	Sensitivity		
	FD-NEAT	ANN	SVM
1	65.6%	68.0%	64.9%
2	83.7%	83.1%	83.1%
3	86.5%	84.9%	85.7%
4	87.5%	87.5%	83.8%

In the LIDC database, radiologists' ratings of some nodule characteristics are assigned to each outlined nodule with diameter exceeding 3 mm. One nodule characteristic, "internal structure," includes the categories "soft tissue," "fluid," "fat," and "air," and another characteristic, "calcification," includes five categories of calcification morphology and distribution, if present. The other characteristics, namely "subtlety," "sphericity," "margin," "lobulation," "spiculation," "texture," and "likelihood of malignancy" are expressed as a single rating on a five-point scale, some of which include descriptive labels for all five points, some have such labels for the two extreme points only, and others also include a label for the middle point.

On the Lung Image Database Consortium research page of the NCI website [33], it is stated that for a subset of cases, inconsistent rating systems were used among the five institutions with regard to the spiculation and lobulation characteristics of nodules > 3 mm. Another statement was released in March 2010 on the NCI website that the inconsistency issue still remains to be

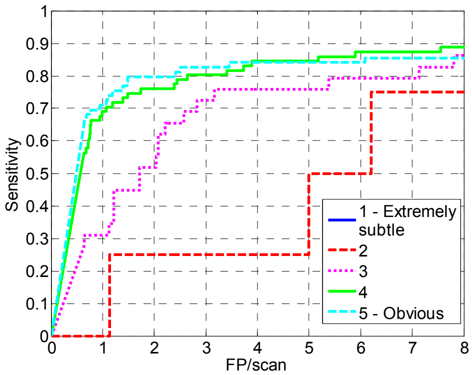

Figure 2: FROC analysis of our CAD system for the FD-NEAT classifier, as a function of nodule subtlety.

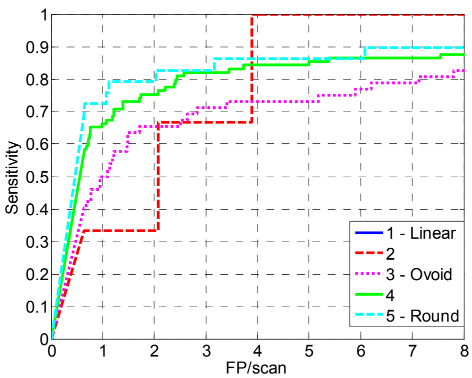

Figure 3: FROC analysis of our CAD system for the FD-NEAT classifier, as a function of nodule sphericity.

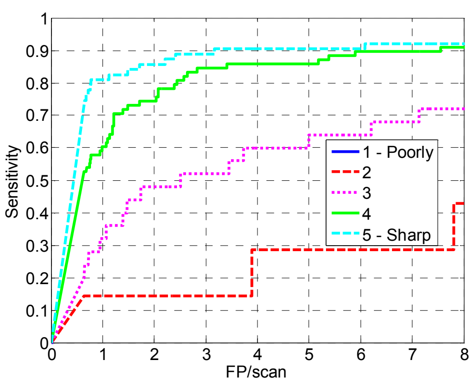

Figure 4: FROC analysis of our CAD system for the FD-NEAT classifier, as a function of nodule margin.

corrected. We thus omitted the analysis on the nodule spiculation and lobulation characteristics here.

We perform the analysis with the FD-NEAT classifier. The analysis is performed on nodules detected at agreement level 2. Previous study showed that nodules at agreement level 1 are detected at much lower sensitivity by our CAD system [1]; however, correct results at agreement level 1 might not be truly indicative of the good performance of a CAD system as the majority of the radiologists did not indicate them as nodules. The FROC curves of the FD-NEAT classifier for the different nodule

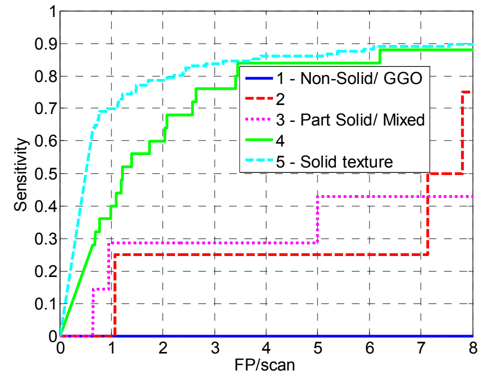

Figure 5: FROC analysis of our CAD system for the FD-NEAT classifier, as a function of nodule internal texture.

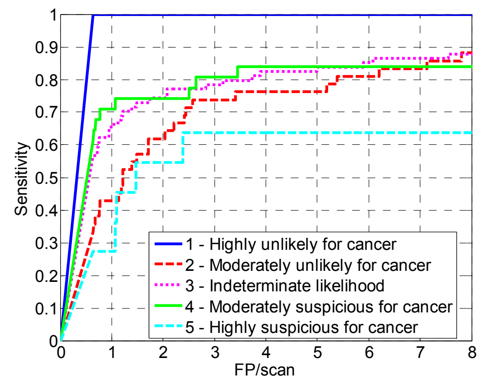

Figure 6: FROC analysis of our CAD system for the FD-NEAT classifier, as a function of nodule likelihood of malignancy.

characteristics are given in Figures 2 to 6. In the analysis, we study the performance of the CAD system as a function of the nodule characteristics. Differences among the radiologist ratings were reconciled by computing the rounded average of the radiologist ratings. We do not retrain the classifier, but simply apply it on the test set. Retraining the classifier for each analyzed nodule characteristic would require the selection and computation of relevant features each time for the nodule characteristic in question. The computational requirement for this task is too high and the practical value of the analysis is limited. Here, we perform an analysis to study the performance of the CAD system in relation to nodule subtlety, sphericity, margin, internal texture, and likelihood of malignancy.

The analysis by nodule subtlety in Figure 2 shows that a strong relation exists between CAD sensitivity and the radiologists' perception of subtlety. The performance of the CAD system with the FD-NEAT classifier is considerably better on nodules that are more obvious (ratings from 3 to 5) than on nodules that are subtle (rating 2 – there are no nodules with rating 1 at agreement level 2). Similar performance trends are reported in [2, 17] for two other CAD systems that were validated on nodules from the LIDC database as a function of the nodule subtlety ratings.

The analysis by nodule sphericity in Figure 3 also shows that a strong relation exists between CAD sensitivity and the radiologists' perception of sphericity. The performance of the CAD system is considerably better on nodules that are more round or spherical (ratings 4 and 5) than on linear or ovoid nodules (ratings 2 and 3 – there are no nodules with rating 1 at agreement level 2). The FROC curve obtained for the nodules with sphericity rating 2 has an unusual shape as there are only three nodules with this rating detected at agreement level 2 in the independent test set.

In Figure 4, the analysis by nodule margin shows that nodules with more sharply-defined margins are detected at much higher sensitivities than poorly-defined ones. The sensitivity results of our CAD system on the test set, as a function of the nodule internal structure ratings, are shown in Figure 5. The CAD system's performance is highest on solid textured nodules and lowest on non-solid nodules. At agreement level 2, there is only one nodule with rating 1 in the test set, which was not detected by the system at FP rates less than or equal to 8 FP/scan. Ohta et al. [34] examine 87 resected lung specimens that show Ground Glass Opacity (GGO) texture on CT, including 47 pure ground-glass lesions. They report that the frequency of invasion of metastasis is low in pure GGOs.

Figure 6 displays the performance of our CAD system on the independent test set as a function of nodule likelihood of malignancy. This characteristic is especially subjective, since the radiologists were not provided with any clinical information about the patients; as a general guide, likelihood of malignancy was rated under the assumption of a 60-year-old male smoker [16]. The results show that the CAD system's performance on nodules with rating 1 is generally the highest, followed by nodules with ratings 4 and 3, respectively. The performance on nodules with ratings 2 and 5 are generally the lowest.

The results of the analysis based on malignancy ratings might be rather surprising. Why would performance on nodules with ratings 4 and 5 be generally outperformed by rating 1? Most probably, this is caused by the specific choice for the 45 input features of the classifier, which are selected to solve the task of nodule detection and not nodule classification (i.e. differentiating between benign and malignant nodules). Other features will be more helpful for nodule classification, e.g. those that measure change in size or growth rate of a nodule [35] and features that measure nodule irregularity [36].

5. ANALYSIS OF NETWORK COMPLEXITY EVOLVED BY FD-NEAT

FD-NEAT is based on the NEAT algorithm that starts minimally and with all input nodes connected to the output nodes, and adds structure (nodes and connections) if they improve network performance. Since FD-NEAT starts minimally, and simultaneously enables the pruning or deletion of initial input connections, it is expected that the evolved networks have a simpler structure in terms of the number of hidden nodes or the number of connections or both. An analysis was performed on the networks evolved by FD-NEAT on the training set of the lungs data. Figures 7 to 9 show the graphs of the maximum fitness, the number of connected inputs and the number of connections of the best network of each generation averaged over 10 runs, each of which ran for 200 generations. The error bars are symmetrical, and are two standard deviation units in length.

From Figure 7, it can be observed that the maximum fitness increases with the generation number: A rapid increase is observed from generation 1 to 10 followed by a more gradual increase from generation 11 onwards. In the graph of connected

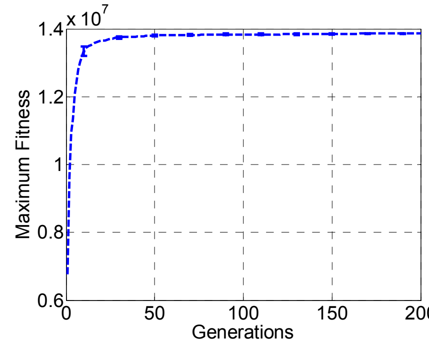

Figure 7: Graph of the maximum fitness of the best network from each generation, averaged over all 10 runs.

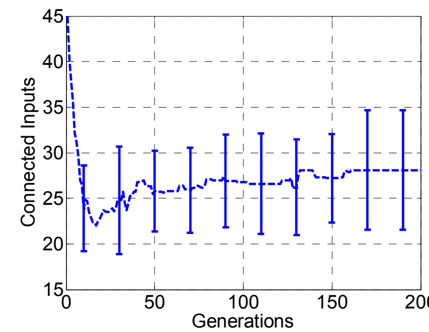

Figure 8: Graph of the number of inputs with at least one connection emerging from them in the best network of each generation, averaged over all 10 runs.

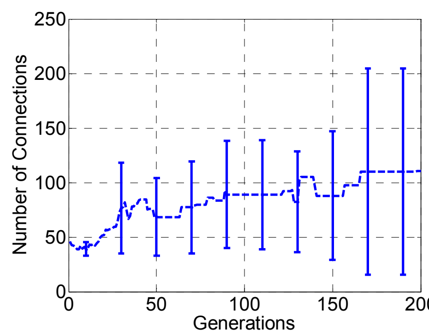

Figure 9: Graph of the number of connections in the best network of each generation, averaged over all 10 runs.

inputs in Figure 8, a considerable drop in the number of connected inputs from 45 (at generation 1) to 22 (at generation 17) is initially observed. After generation 17, there is a general increase in the number of connected inputs and the graph reaches a steady value around generation 160. The same trend is observed in the graph of network sizes in Figure 9 whereby an initial drop is observed in the number of connections from 46 to 39 after which the graph generally increases and reaches a plateau around generation 170.

From Figures 8 and 9, it can be observed that the standard deviation intervals of the input connections and especially of the network size are very big. The standard deviation intervals for the maximum fitness graph of Figure 7 are comparatively much smaller. The results indicate that the network structure evolved by FD-NEAT can vary greatly and hence, it might be difficult to make a specific or targeted comparison with the fixed-topology ANNs.

It can also be observed that the standard deviation intervals of the input connections and of the network size generally increases with the number of generations. This might be due to the fact that different features are highly correlated. If this is the case, the same fitness can be obtained by selecting a single feature that is highly correlated with the classification and yet is uncorrelated to other features compared with selecting a feature subset of a few features that are highly correlated to each other, but are not so correlated with the classification task. Hence, bigger variation is to be expected in the number of inputs and network size if the experiments are extended further to higher generation numbers.

From Figure 9, it is observed that the number of connections averaged over all 10 runs reaches a plateau value of 110 after ~170 generations. The number of connections in a fully-connected, cycle-free 11-hidden-node ANN = (45×11 input connections) + (11×1 output connections) + 11 hidden layer bias connections + 1 output bias connection = 518. If we compare this value with the network size attained by FD-NEAT after 170 generations (i.e. 110±95 connections), we see that the sizes of the networks evolved by FD-NEAT are still smaller than in the case of a fixed ANN (The FD-NEAT algorithm written in Matlab® also evolves cycle-free networks so valid comparisons with the fully-connected, cycle-free fixed-topology ANN can be made). The average number of nodes evolved by FD-NEAT (i.e. the sum of the bias, connected input, output, and hidden nodes), averaged over the 10 runs = 55.9±32.2. The total number of nodes of the fully-connected 11-hidden-node ANN is 57.

In summary, from the results of network size, it is observed that FD-NEAT produces networks that have much fewer connections compared to the fixed-topology ANN. The average number of nodes evolved by FD-NEAT is slightly lower than the ANN; however, the standard deviation of the number of nodes evolved by FD-NEAT is too high to make a conclusive comparison with the ANN.

6. CONCLUSIONS

We performed an analysis of our CAD system with the feature-deselective neuroevolution classifier, FD-NEAT, as a function of various nodule characteristics defined in the LIDC database. We also performed an analysis of network complexity based on maximum fitness, connected inputs and network size on the networks evolved by FD-NEAT for the lung nodule detection experiments. The analysis on nodule characteristics shows that the sensitivity results of FD-NEAT relates well with the radiologists' perception of nodule subtlety, sphericity, margin, and internal texture. A strong relation of the FD-NEAT classifier with the radiologists' perception of nodule likelihood of malignancy is not present, as FD-NEAT was trained on features for lung nodule detection, and not nodule classification (benign/ malignant differentiation). The analysis on network complexity shows that FD-NEAT evolves networks that are less complex than the ANN in terms of number of connections, but not always in terms of number of nodes.

7. ACKNOWLEDGMENTS

The authors acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health and their critical role in the creation of the free publicly available LIDC/IDRI Database used in this study. Image data for this research were obtained from The Cancer Imaging Archive (http://cancerimagingarchive.net/) sponsored by the Cancer Imaging Program, DCTD/NCI/NIH.

8. APPENDIX

This Appendix describes the FD-NEAT parameters used for experiments in this paper. The list is not exhaustive; the reader is referred to [1] for a complete list of all the parameter values used. Each population has 200 networks. The coefficients for measuring compatibility are c_1 (coefficient to determine the importance of excess genes in measuring compatibility) = 1.0, c_2 (coefficient to determine the importance of disjoint genes in measuring compatibility) = 1.0, and c_3 (coefficient to determine the importance of average weight difference in measuring compatibility) = 0.3. The initial compatibility distance for speciation, C_t is 8.0. However, because the population dynamics can be unpredictable over the course of evolution, we assign a target of 10 species. If the number of species exceeds 10, C_t is increased by 4.0 to reduce the number of species. Conversely, if the number of species is less than 10, C_t is decreased by 4.0 to increase the number of species. The champion of each species with more than five networks is copied unchanged into the next generation. There is a 10% chance that an inherited gene is re-enabled in the offspring if it is inherited disabled. Conversely, there is a 15% chance that an inherited gene is disabled in the offspring if it is inherited enabled – This probability only applies to the initial input connections, and is the principle of selecting relevant features for FD-NEAT. The probability that recurrent connections are formed is put to zero. The interspecies mating rate is only 5%. The probability of adding a new node is set initially to 0.5, and the probability of adding a new link or connection to 0.8. After 20 generations from the start of evolution, the add node probability is changed to 0.05, and the add link probability to 0.9. After 45 generations, the add link probability is modified to be 0.1 whereas the add node probability is maintained at 0.05.

9. REFERENCES

[1] Tan, M., Deklerck, R., Jansen, B., Bister, M., and Cornelis, J. 2011. A novel computer-aided lung nodule detection system for CT images. *Med. Phys.* 38, 10 (Oct. 2011), 5630-5645.

[2] Messay, T., Hardie, R. C. and Rogers, S. K. 2010. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med. Image Anal.* 14, 3 (Jun. 2010), 390-406.

[3] Li, Q., Sone, S. and Doi, K. 2003. Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans. *Med. Phys.* 30, 8 (Aug. 2003), 2040-2051.

[4] American Cancer Society. 2009. Cancer Facts and Figures.

[5] Fry, W. A., Menck, H. R. and Winchester, D. P. 1996. The National cancer data base report on lung cancer. *Cancer*. 77, 9 (May 1996), 1947-1955.

[6] Flehinger, B. J., Kimmel, M. and Melamed, M. R. 1992. The effect of surgical treatment on survival from early lung cancer. Implications for screening. *Chest.* 101, 4 (Apr. 1992), 1013-1018.

[7] Martini, N., Bains, M. S., Burt, M. E., Zakowski, M. F., McCormack, P., Rusch, V. W., and Ginsberg, R. J. 1995. Incidence of local recurrence and second primary tumors in

resected stage I lung cancer. *The Journal of Thoracic and Cardiovascular Surgery*. 109, 1 (Jan. 1995), 120-129.

[8] Henschke, C. I., McCauley, D. I., Yankelevitz, D. F., Naidich, D. P., McGuinness, G., Miettinen, O. S., Libby, D. M., Pasmantier, M. W., Koizumi, J., Altorki, N. K., and Smith, J. P. 1999. Early Lung Cancer Action Project: Overall design and findings from baseline screening. *Lancet*. 354, 9173 (Jul. 1999), 99-105.

[9] Swensen, S. J., Jett, J. R., Hartman, T. E., Midthun, D. E., Sloan, J. A., Sykes, A.-M., Aughenbaugh, G. L., and Clemens, M. A. 2003. Lung cancer screening with CT: Mayo Clinic experience. *Radiology*. 226, 3 (Mar. 2003), 756-761.

[10] Kaneko, M., Eguchi, K., Ohmatsu, H., Kakinuma, R., Naruke, T., Suemasu, K., and Moriyama, N. 1996. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology*. 201, 3 (Dec. 1996), 798-802.

[11] Xu, D. M., Gietema, H., Koning, H. d., Vernhout, R., Nackaerts, K., Prokop, M., Weenink, C., Lammers, J.-W., Groen, H., Oudkerk, M., and Klaveren, R. v. 2006. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer*, 54, 2 (Nov. 2006), 177-184.

[12] Barnes, E. 2010. NLST results show drop in lung cancer deaths with CT screening. AuntMinnie.com, (Nov. 2010). http://www.auntminnie.com/index.aspx?sec=sup&sub=cto&pag=dis&itemid=92678&wf=4035. Last accessed: February 2012.

[13] National Lung Screening Trial Research Team. 2011. The National Lung Screening Trial: Overview and study design. *Radiology*. 258, 1 (Jan. 2011), 243-253.

[14] Tan, M., Hartley, M., Bister, M., and Deklerck, R. 2009. Automated feature selection in neuroevolution. *Evol. Intel.* 4, 1 (Feb. 2009), 271-292.

[15] Stanley, K. O. and Miikkulainen, R. 2002. Evolving neural networks through augmenting topologies. *Evol. Comput.* 10, 2, 99-127.

[16] Armato, S. G. III, McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., van Beek, E. J. R., Yankelevitz, D., et al. 2011. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med. Phys.* 38, 2 (Feb. 2011), 915-931.

[17] Golosio, B., Masala, G.L., Piccioli, A., Oliva, P., Carpinelli, M., Cataldo, R., Cerello, P., De Carlo, F., Falaschi, F., Fantacci, M.E., Gargano, G., Kasae, P., and Torsello, M., 2009. A novel multithreshold method for nodule detection in lung CT. *Med. Phys.* 36, 8 (Aug. 2009), 3607-3618.

[18] Opfer, R., Wiemker, R., 2007. Performance analysis for computer aided lung nodule detection on LIDC data. In *Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment: Proceedings of SPIE.* 6515, 65151C

[19] Boroczky, L., Zhao, L. and Lee, K. P. 2006. Feature subset selection for improving the performance of false positive reduction in lung nodule CAD. *IEEE Trans. Inf. Technol. Biomed.* 10, 3 (Jul. 2006), 504-511.

[20] Suarez-Cuenca, J. J., Tahoces, P. G., Souto, M., Lado, M. J., Remy-Jardin, M., Remy, J., and Vidal, J. J. 2009. Application of the iris filter for automatic detection of pulmonary nodules on computed tomography images. *Comput. Biol. Med.* 39, 10 (Oct. 2009), 921-933.

[21] Skalak, D. B. 1994. Prototype and feature selection by sampling and random mutation hill-climbing algorithms. In: *Proceedings of the 11th International Conference on Machine Learning.* Morgan Kaufmann, New Brunswick, 293-301.

[22] Langley, P. and Sage, S. 1994. Oblivious decision trees and abstract cases. In: *Working notes of the AAAI-94 Workshop on Case-Based Reasoning.* AAAI Press, Seattle, 113-117.

[23] Mayr, C. 2003. NEAT Matlab. Last accessed: February 2012. http://nn.cs.utexas.edu/?neatmatlab.

[24] Chang, C.-C. and Lin, C.-J. 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 3, Article 27 (Apr. 2011), 1-27. Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[25] Hsu, C.-W., Chang, C.-C. and Lin, C.-J. 2009. *A practical guide to support vector classification*. Technical report, National Taiwan University, Taipei 106, Taiwan.

[26] Marquardt, D. 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math.* 11, 2 (Jun. 1963), 431-441.

[27] Hagan, M. T. and Menhaj, M. 1999. Training feed-forward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw*. 5, 6 (Nov. 1994), 989-993.

[28] Timin, M. E. 1995. The robot auto racing simulator. Last accessed: February 2012. http://rars.sourceforge.net.

[29] Salden, A. H., Romeny, B. M. T., Viergever, M. A., Florack, L., and Koenderink, J. J. 1992. *Differential geometric description of 3D scalar images.* Technical Report. 3D Computer Vision, Utrecht University Hospital, The Netherlands.

[30] Romeny, B. M. t. H. 2003. *Front-End Vision & Multi-Scale Image Analysis*. Springer Verlag, Berlin.

[31] Kohonen, T. 1997. *Self-Organizing Maps*. 2nd ed. Springer, Heidelberg.

[32] Kohonen, T. 1988. *Self-Organization and Associative Memory*. 3rd ed. Springer Verlag, Berlin, Heidelberg.

[33] National Cancer Institute. 2012. Lung Image Database Consortium. Last accessed: February 2012. https://wiki.cancerimagingarchive.net/display/Public/Lung+Image+Database+Consortium.

[34] Ohta, Y., Shimizu, Y., Kobayashi, T., Matsui, O., Minato, H., Matsumoto, I., Watanabe, G., 2006. Pathologic and biological assessment of lung tumors showing ground-glass opacity. *Ann. Thorac. Surg.* 81, 4 (Apr. 2006), 1194–1197.

[35] Reeves, A. P., Chan, A. B., Yankelevitz, D. F., Henschke, C. I., Kressler, B., Kostis, W. J., 2006. On measuring the change in size of pulmonary nodules. *IEEE Trans. Med. Imaging.* 25, 4 (Apr. 2006), 435-450.

[36] Aoyama, M., Li, Q., Katsuragawa, S., MacMahon, H., Doi, K., 2002. Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images. *Med. Phys*. 29, 5 (May 2002), 701-708.