Lung Cancer Identification by an Electronic Nose based on an Array of MOS Sensors

Rossella Blatt\textsuperscript{1}, Andrea Bonarini\textsuperscript{1}, Elisa Calabrò\textsuperscript{2}
Matteo Della Torre\textsuperscript{3}, Matteo Matteucci\textsuperscript{1} and Ugo Pastorino\textsuperscript{2}

Speaker: Rossella Blatt
blatt@elet.polimi.it

\textsuperscript{1}Politecnico di Milano, Department of Electronics and Information, Milan, Italy
\textsuperscript{2}Istituto Nazionale Tumori of Milan, Toracic Surgery Department, Milan, Italy
\textsuperscript{3}SACMI Imola S.C., Automation & Inspection Systems, Imola (BO), Italy
Outline

- **Objective:** *Lung Cancer* diagnosis classifying the *Olfactory Signal* acquired by an *Electronic Nose*
- Motivation
- Functioning of the Electronic Nose
- Classification of volunteers' breath
- Results and comparison with current diagnostic techniques
- Further directions of research
Outline

- Objective: Lung Cancer diagnosis classifying the Olfactory Signal acquired by an Electronic Nose
- Motivation
  - Functioning of the Electronic Nose
  - Classification of volunteers' breath
  - Results and comparison with current diagnostic techniques
  - Further directions of research
Outline

- **Objective:** Lung Cancer diagnosis classifying the Olfactory Signal acquired by an Electronic Nose
- Motivation
- Functioning of the Electronic Nose
- Classification of volunteers' breath
- Results and comparison with current diagnostic techniques
- Further directions of research
Objective: Lung Cancer diagnosis classifying the Olfactory Signal acquired by an Electronic Nose

Motivation

Functioning of the Electronic Nose

Classification of volunteers' breath

Results and comparison with current diagnostic techniques

Further directions of research
Objective: Lung Cancer diagnosis classifying the Olfactory Signal acquired by an Electronic Nose

Motivation

Functioning of the Electronic Nose

Classification of volunteers' breath

Results and comparison with current diagnostic techniques

Further directions of research
Objective: Lung Cancer diagnosis classifying the Olfactory Signal acquired by an Electronic Nose

Motivation

Functioning of the Electronic Nose

Classification of volunteers' breath

Results and comparison with current diagnostic techniques

Further directions of research
Lung cancer causes more than 160,000 deaths a year in the United States—more than any other cancer

Once lung cancer is detected the probability of surviving, after 5 years of therapy, is 14%; the survival rate increases to 48% if the cancer is discovered in its earliest stage

Current diagnostic techniques are invasive, very expensive, have a high risk of complications and a not so good performance

Efforts at early detection and treatment have been frustrating to date and hence the overall prognosis remains poor

NECESSITY OF A NON INVASIVE, MORE ACCURATE AND CHEAPER DIAGNOSTIC TECHNIQUE, ABLE TO IDENTIFY THE PRESENCE OF LUNG CANCER IN ITS EARLY STAGE
Motivation

- Lung cancer causes more than 160,000 deaths a year in the United States--more than any other cancer
- Once lung cancer is detected the probability of surviving, after 5 years of therapy, is 14%; the survival rate increases to 48% if the cancer is discovered in its earliest stage
- Current diagnostic techniques are invasive, very expensive, have a high risk of complications and a not so good performance
- Efforts at early detection and treatment have been frustrating to date and hence the overall prognosis remains poor

NECESSITY OF A NON INVASIVE, MORE ACCURATE AND CHEAPER DIAGNOSTIC TECHNIQUE, ABLE TO IDENTIFY THE PRESENCE OF LUNG CANCER IN ITS EARLY STAGE
Motivation

- Lung cancer causes more than 160,000 deaths a year in the United States--more than any other cancer
- Once lung cancer is detected the probability of surviving, after 5 years of therapy, is 14%; the survival rate increases to 48% if the cancer is discovered in its earliest stage
- Current diagnostic techniques are invasive, very expensive, have a high risk of complications and a not so good performance
- Efforts at early detection and treatment have been frustrating to date and hence the overall prognosis remains poor

NECESSITY OF A NON INVASIVE, MORE ACCURATE AND CHEAPER DIAGNOSTIC TECHNIQUE, ABLE TO IDENTIFY THE PRESENCE OF LUNG CANCER IN ITS EARLY STAGE
Motivation

- Lung cancer causes more than 160,000 deaths a year in the United States--more than any other cancer
- Once lung cancer is detected the probability of surviving, after 5 years of therapy, is 14%; the survival rate increases to 48% if the cancer is discovered in its earliest stage
- Current diagnostic techniques are invasive, very expensive, have a high risk of complications and a not so good performance
- Efforts at early detection and treatment have been frustrating to date and hence the overall prognosis remains poor

NECESSITY OF A NON INVASIVE, MORE ACCURATE AND CHEAPER DIAGNOSTIC TECHNIQUE, ABLE TO IDENTIFY THE PRESENCE OF LUNG CANCER IN ITS EARLY STAGE
Motivation

- **Fundamental Principle of Clinical Chemistry**: “Every pathology changes people chemical composition, modifying the concentration of some chemicals in the human body”
  - In the medical field, clinicians have always considered odor as a fundamental information for the diagnosis of several diseases

- It has been demonstrated (Gordon et al, 1985) that the presence of lung cancer alters the percentage of some volatile organic compounds (VOCs) present in human breath
  - These VOCs can be considered as lung cancer markers and thus used to diagnose it
Motivation

- **Fundamental Principle of Clinical Chemistry:** “Every pathology changes people chemical composition, modifying the concentration of some chemicals in the human body”
  - In the medical field, clinicians have always considered odor as a fundamental information for the diagnosis of several diseases

- It has been demonstrated (Gordon et al, 1985) that the presence of lung cancer alters the percentage of some volatile organic compounds (VOCs) present in human breath
  - These VOCs can be considered as lung cancer markers and thus used to diagnose it
**Motivation**

- **Fundamental Principle of Clinical Chemistry:** “Every pathology changes people chemical composition, modifying the concentration of some chemicals in the human body”
  - In the medical field, clinicians have always considered odor as a fundamental information for the diagnosis of several diseases
- It has been demonstrated (Gordon et al, 1985) that the presence of lung cancer alters the percentage of some volatile organic compounds (VOCs) present in human breath
  - These VOCs can be considered as lung cancer markers and thus used to diagnose it
An electronic nose is an instrument able to **acquire, detect and analyse the olfactory signal**

- It is composed of an **array of non specific electronic devices** (sensors) able to **convert a physical or chemical information into an electrical signal**
  - It is non specific because it does not look for particular compounds in the analyzed substance, but for different **patterns**
  - Each sensor reacts in a different way to the analyzed substance providing **multidimensional data** that can be considered as an **olfactory blueprint** of the substance itself
Electronic Nose

- An electronic nose is an instrument able to acquire, detect and analyse the olfactory signal.

- It is composed of an array of non specific electronic devices (sensors) able to convert a physical or chemical information into an electrical signal.

  - It is non specific because it does not look for particular compounds in the analyzed substance, but for different patterns.
  - Each sensor reacts in a different way to the analyzed substance providing multidimensional data that can be considered as an olfactory blueprint of the substance itself.
Electronic Nose

- An electronic nose is an instrument able to acquire, detect and analyse the olfactory signal
- It is composed of an array of non specific electronic devices (sensors) able to convert a physical or chemical information into an electrical signal
  - It is non specific because it does not look for particular compounds in the analyzed substance, but for different patterns
  - Each sensor reacts in a different way to the analyzed substance providing multidimensional data that can be considered as an olfactory blueprint of the substance itself
Electronic Nose

- An electronic nose is an instrument able to acquire, detect and analyse the olfactory signal.
- It is composed of an array of non specific electronic devices (sensors) able to convert a physical or chemical information into an electrical signal.
  - It is non specific because it does not look for particular compounds in the analyzed substance, but for different patterns.
  - Each sensor reacts in a different way to the analyzed substance providing multidimensional data that can be considered as an olfactory blueprint of the substance itself.
According to the used pattern analysis algorithm, the output of an electronic nose can be:

- the **detection** of a specific substance
- an estimate of the **concentration** of the odor
- some particular characteristic of the odor that allows to associate it to a particular **class**
Electronic Nose's Output

- According to the used pattern analysis algorithm, the output of an electronic nose can be:
  - the **detection** of a specific substance
  - an estimate of the **concentration** of the odor
  - some particular characteristic of the odor that allows to associate it to a particular class
According to the used pattern analysis algorithm, the output of an electronic nose can be:

- the detection of a specific substance
- an estimate of the concentration of the odor
- some particular characteristic of the odor that allows to associate it to a particular class
Electronic Nose's Output

- According to the used pattern analysis algorithm, the output of an electronic nose can be:
  - the **detection** of a specific substance
  - an estimate of the **concentration** of the odor
  - some particular characteristic of the odor that allows to associate it to a particular **class**
Electronic Nose

1. Signal Acquisition
   - Acquisition is done through a sensor array that measures a given physical or chemical quantity and convert it into an electrical signal

2. Signal Processing
   - Preprocessing: aimed to reduce the impact of noise
   - Dimensionality Reduction: reduce the dimensionality of the problem, enhancing classification performance

3. Classification and Validation
   - Classification between the two classes “healthy” and “sick”
1. Signal Acquisition
   - Acquisition is done through a sensor array that measures a given physical or chemical quantity and convert it into an electrical signal

2. Signal Processing
   - Preprocessing: aimed to reduce the impact of noise
   - Dimensionality Reduction: reduce the dimensionality of the problem, enhancing classification performance

3. Classification and Validation
   - Classification between the two classes “healthy” and “sick”
1. **Signal Acquisition**
   - Acquisition is done through a sensor array that measures a given physical or chemical quantity and convert it into an electrical signal

2. **Signal Processing**
   - **Preprocessing**: aimed to reduce the impact of noise
   - **Dimensionality Reduction**: reduce the dimensionality of the problem, enhancing classification performance

3. **Classification and Validation**
   - Classification between the two classes “healthy” and “sick”
1. **Signal Acquisition**
   - Acquisition is done through a sensor array that measures a given physical or chemical quantity and convert it into an electrical signal

2. **Signal Processing**
   - *Preprocessing*: aimed to reduce the impact of noise
   - *Dimensionality Reduction*: reduce the dimensionality of the problem, enhancing classification performance

3. **Classification and Validation**
   - Classification between the two classes “healthy” and “sick”
Lung Cancer Identification by an Electronic Nose based on an Array of MOS Sensors

Signal Acquisition

**SIGNS ACQUISITION**
- Gas to Analyse
- Sensors Array
- Acquisition system

**PREPROCESSING & DIMENSIONALITY REDUCTION**
- Baseline Manipulation
- Noise Reduction
- Filter Normalization
- Feature Extraction
- Feature Selection
- PCA - LDA

**CLASSIFICATION**
- k-Nearest Neighbors
- Linear Discriminant Analysis
- Quadratic Discriminant Analysis
- Artificial Neural Networks
Signal Acquisition

- The breath acquisition has been made inviting all volunteers to blow into a nalophan bag of approximately 400cm³

- Then, the air contained in the bag was input into the electronic nose

- We used an array of six MOS sensors that react to gases with a variation of resistance

- The registered signal corresponds to the change of resistance through time produced by gas flow
Signal Acquisition

- The breath acquisition has been made inviting all volunteers to blow into a nalophan bag of approximately 400 cm$^3$.

- Then, the air contained in the bag was input into the electronic nose.
  - We used an array of six MOS sensors that react to gases with a variation of resistance.
  - The registered signal corresponds to the change of resistance through time produced by gas flow.
Signal Acquisition

- The breath acquisition has been made inviting all volunteers to blow into a nalophan bag of approximately 400cm³.

- Then, the air contained in the bag was input into the electronic nose.

- We used an array of six MOS sensors that react to gases with a variation of resistance.

- The registered signal corresponds to the change of resistance through time produced by gas flow.
Signal Acquisition

- The breath acquisition has been made inviting all volunteers to blow into a nalophan bag of approximately 400cm$^3$

- Then, the air contained in the bag was input into the electronic nose

- We used an array of six MOS sensors that react to gases with a variation of resistance

- The registered signal corresponds to the change of resistance through time produced by gas flow
Involved Population

- We analyzed the breath of **101 volunteers**

- For each person we took two measures for a total of **202 measurements** (116 healthy, 86 diseased)
Involved Population

- We analyzed the breath of 101 volunteers

Total population 101 volunteers

Lung Cancer 43

Healthy 58

Primary Lung Cancer 23

Pulmonary Metastasis 20

- For each person we took two measures for a total of 202 measurements (116 healthy, 86 diseased)
Involved Population

- We analyzed the breath of **101 volunteers**

- For each person we took two measures for a total of **202 measurements** (116 healthy, 86 diseased)
Involved Population

- We analyzed the breath of 101 volunteers

- For each person we took two measures for a total of 202 measurements (116 healthy, 86 diseased)
Lung Cancer Identification by an Electronic Nose based on an Array of MOS Sensors

Pre-processing & Dimensionality Reduction

Gas to Analyse → Sensors Array → Acquisition system →

PREPROCESSING & DIMENSIONALITY REDUCTION

Baseline Manipulation
Noise Reduction
Filter Normalization
Feature Extraction
Feature Selection
PCA - LDA

CLASSIFICATION

k-Nearest Neighbors
Linear Discriminant
Quadratic Discriminant
Artificial Neural Networks
Signal pre-processing

- Manipulation of the **baseline**: transformation of the sensor response w.r.t. its baseline for the purpose of **drift compensation**
- Reduction of **humidity** effects
- **Normalization**: compensation for the scale difference among the six sensors
  - Each sensor has been forced to have zero mean and variance equal to 1
Signal pre-processing

- Manipulation of the baseline: transformation of the sensor response w.r.t. its baseline for the purpose of drift compensation
- Reduction of humidity effects
- Normalization: compensation for the scale difference among the six sensors
  - Each sensor has been forced to have zero mean and variance equal to 1
Signal pre-processing

- Manipulation of the **baseline**: transformation of the sensor response w.r.t. its baseline for the purpose of **drift compensation**
- Reduction of **humidity** effects
- **Normalization**: compensation for the scale difference among the six sensors
  - Each sensor has been forced to have zero mean and variance equal to 1
Dimensionality Reduction

- Feature Selection
  - Non-parametric test of Mann-Whitney-Wilcoxon
  - Scatter Plot
  - MANOVA

- Feature Extraction
Dimensionality Reduction

- Feature Selection
  - Non-parametric test of Mann-Whitney-Wilcoxon
  - Scatter Plot
  - MANOVA

- Feature Extraction
  - Non Parametric Linear Discriminant Analysis NPLDA (Fukunaga, 1983)
    - A generalization of Fisher's LDA
    - It removes the unimodal gaussian assumption by computing the between scatter matrix $S_b$ using the $k$-NN rule
  - Best projection: 1st NPLDA component
Dimensionality Reduction

- Feature Selection
  - Non-parametric test of Mann-Whitney-Wilcoxon
  - Scatter Plot
  - MANOVA

- Feature Extraction
  - Non Parametric Linear Discriminant Analysis NPLDA (Fukunaga, 1983)
    - A generalization of Fisher's LDA
    - It removes the unimodal gaussian assumption by computing the between scatter matrix $S_b$ using the $k$-NN rule
  - Best projection: 1st NPLDA component
Dimensionality Reduction

- **Feature Selection**
  - Test of Mann-Whitney-Wilcoxon
  - Scatter Plot
  - MANOVA

- **Feature Extraction**
  - NPLDA (Fukunaga, 1983)
    - A generalization of Fisher's LDA
    - It removes the unimodal gaussian assumption by computing the between scatter matrix $S_b$ using the $k$-NN rule
  - Best projection: 1st NPLDA component
Lung Cancer Identification by an Electronic Nose based on an Array of MOS Sensors
Classification

- Different families of classifiers:
  - Nearest Neighbor Classifiers (k-NN)
    - Classic k-NN
    - Modified k-NN --> $k=$number of neighbors belonging all to the same class
    - Fuzzy $k$-Nearest Neighbors --> assigns a class membership function to each training and test samples
  - Discriminant Functions Classifiers
    - Linear
    - Quadratic
  - Artificial Neural Network
    - Feedforward Neural Network with one hidden layer
Classification

- Different families of classifiers:
  - Nearest Neighbor Classifiers (k-NN)
    - Classic k-NN
    - Modified k-NN --> $k =$ number of neighbors belonging all to the same class
    - Fuzzy $k$-Nearest Neighbors --> assigns a class membership function to each training and test samples
  - Discriminant Functions Classifiers
    - Linear
    - Quadratic
  - Artificial Neural Network
    - Feedforward Neural Network with one hidden layer
Classification

- Different families of classifiers:
  - Nearest Neighbor Classifiers (k-NN)
    - Classic k-NN
    - Modified k-NN --> $k=$ number of neighbors belonging all to the same class
    - Fuzzy $k$-Nearest Neighbors --> assigns a class membership function to each training and test samples
  - Discriminant Functions Classifiers
    - Linear
    - Quadratic
  - Artificial Neural Network
    - Feedforward Neural Network with one hidden layer
Classification

- Different families of classifiers:
  - Nearest Neighbor Classifiers (k-NN)
    - Classic k-NN
    - Modified k-NN $\rightarrow k=$ number of neighbors belonging all to the same class
    - Fuzzy $k$-Nearest Neighbors $\rightarrow$ assigns a class membership function to each training and test samples
  - Discriminant Functions Classifiers
    - Linear
    - Quadratic
  - Artificial Neural Network
    - Feedforward Neural Network with one hidden layer
Classification

- Different families of classifiers:
  - Nearest Neighbor Classifiers (k-NN)
    - Classic k-NN
    - Modified k-NN --> $k =$ number of neighbors belonging all to the same class
    - Fuzzy $k$-Nearest Neighbors --> assigns a class membership function to each training and test samples
  - Discriminant Functions Classifiers
    - Linear
    - Quadratic
  - Artificial Neural Network
    - Feedforward Neural Network with one hidden layer
Results

- Performance has been evaluated through **confusion matrix** and the corresponding **performance indexes** (CI = 95%)
- **Cross-validation**: *modified* leave-one-out
- **We considered different values for** \( k \) (\( k = 1, 3, 5, 9, 101 \))

<table>
<thead>
<tr>
<th>Classifier</th>
<th>NER</th>
<th>TPR</th>
<th>TNR</th>
<th>( \text{PREC}_{POS} )</th>
<th>( \text{PREC}_{NEG} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic 9-NN</td>
<td>90.1%</td>
<td>89.5%</td>
<td>90.5%</td>
<td>87.5%</td>
<td>92.1%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[85.7-94.5]</td>
<td>[85.3-93.8]</td>
<td>[86.0-95.0]</td>
<td>[81.6-93.4]</td>
<td>[86.8-97.4]</td>
</tr>
<tr>
<td>Modified 9-NN</td>
<td>91.1%</td>
<td>91.9%</td>
<td>90.5%</td>
<td>87.8%</td>
<td>93.7%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[86.8-95.4]</td>
<td>[87.9-95.9]</td>
<td>[86.0-95.0]</td>
<td>[81.9-93.7]</td>
<td>[89.1-98.4]</td>
</tr>
<tr>
<td>Fuzzy ( k )-NN</td>
<td>92.6%</td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
<tr>
<td>LD</td>
<td>89.6%</td>
<td>96.5%</td>
<td>84.5%</td>
<td>82.2%</td>
<td>97.0%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[85.0-94.2]</td>
<td>[93.7-99.3]</td>
<td>[79.1-89.9]</td>
<td>[75.2-89.1]</td>
<td>[93.9-100]</td>
</tr>
<tr>
<td>QD</td>
<td>92.6%</td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
<tr>
<td>ANN</td>
<td>91.6%</td>
<td>91.9%</td>
<td>91.3793%</td>
<td>88.8%</td>
<td>93.8%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[87.4-95.8]</td>
<td>[87.9-95.9]</td>
<td>[87.0-95.8]</td>
<td>[84.1-93.4]</td>
<td>[88.2-99.4]</td>
</tr>
</tbody>
</table>
**Results**

- Performance has been evaluated through **confusion matrix** and the corresponding **performance indexes** (CI = 95%)
- **Cross-validation:** *modified* leave-one-out
- We considered **different values for k** ($k=1,3,5,9,101$)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>NER</th>
<th>TPR</th>
<th>TNR</th>
<th>PREC$_{POS}$</th>
<th>PREC$_{NEG}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic 9-NN</td>
<td>90.1%</td>
<td>89.5%</td>
<td>90.5%</td>
<td>87.5%</td>
<td>92.1%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[85.7-94.5]</td>
<td>[85.3-93.8]</td>
<td>[86.0-95.0]</td>
<td>[81.6-93.4]</td>
<td>[86.8-97.4]</td>
</tr>
<tr>
<td>Modified 9-NN</td>
<td>91.1%</td>
<td>91.9%</td>
<td>90.5%</td>
<td>87.8%</td>
<td>93.7%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[86.8-95.4]</td>
<td>[87.9-95.9]</td>
<td>[86.0-95.0]</td>
<td>[81.9-93.7]</td>
<td>[89.1-98.4]</td>
</tr>
<tr>
<td>Fuzzy k-NN</td>
<td>92.6%</td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
<tr>
<td>LD</td>
<td>89.6%</td>
<td><strong>96.5%</strong></td>
<td>84.5%</td>
<td>82.2%</td>
<td><strong>97.0%</strong></td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[85.0-94.2]</td>
<td>[93.7-99.3]</td>
<td>[79.1-89.9]</td>
<td>[75.2-89.1]</td>
<td>[93.9-100]</td>
</tr>
<tr>
<td>QD</td>
<td>92.6%</td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
<tr>
<td>ANN</td>
<td>91.6%</td>
<td>91.9%</td>
<td><strong>91.3793%</strong></td>
<td><strong>88.8%</strong></td>
<td>93.8%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[87.4-95.8]</td>
<td>[87.9-95.9]</td>
<td>[87.0-95.8]</td>
<td>[84.1-93.4]</td>
<td>[88.2-99.4]</td>
</tr>
</tbody>
</table>
## Results

- Performance has been evaluated through **confusion matrix** and the corresponding **performance indexes** (CI = 95%)
- **Cross-validation**: *modified* leave-one-out
- We considered **different values for \( k \) (\( k=1,3,5,9,101 \))**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>NER</th>
<th>TPR</th>
<th>TNR</th>
<th>PREC&lt;sub&gt;POS&lt;/sub&gt;</th>
<th>PREC&lt;sub&gt;NEG&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic 9-NN</td>
<td>90.1%</td>
<td>89.5%</td>
<td>90.5%</td>
<td>87.5%</td>
<td>92.1%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[85.7-94.5]</td>
<td>[85.3-93.8]</td>
<td>[86.0-95.0]</td>
<td>[81.6-93.4]</td>
<td>[86.8-97.4]</td>
</tr>
<tr>
<td>Modified 9-NN</td>
<td>91.1%</td>
<td>91.9%</td>
<td>90.5%</td>
<td>87.8%</td>
<td>93.7%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[86.8-95.4]</td>
<td>[87.9-95.9]</td>
<td>[86.0-95.0]</td>
<td>[81.9-93.7]</td>
<td>[89.1-98.4]</td>
</tr>
<tr>
<td>Fuzzy ( k )-NN</td>
<td><strong>92.6%</strong></td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td><strong>96.3%</strong></td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
<tr>
<td>LD</td>
<td>89.6%</td>
<td><strong>96.5%</strong></td>
<td>84.5%</td>
<td>82.2%</td>
<td><strong>97.0%</strong></td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[85.0-94.2]</td>
<td>[93.7-99.3]</td>
<td>[79.1-89.9]</td>
<td>[75.2-89.1]</td>
<td>[93.9-100]</td>
</tr>
<tr>
<td>QD</td>
<td><strong>92.6%</strong></td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td><strong>96.3%</strong></td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
<tr>
<td>ANN</td>
<td>91.6%</td>
<td>91.9%</td>
<td><strong>91.3793%</strong></td>
<td>88.8%</td>
<td>93.8%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[87.4-95.8]</td>
<td>[87.9-95.9]</td>
<td>[87.0-95.8]</td>
<td>[84.1-93.4]</td>
<td>[88.2-99.4]</td>
</tr>
</tbody>
</table>
Results

- Performance has been evaluated through **confusion matrix** and the corresponding **performance indexes** (CI=95%)
- **Cross-validation**: *modified* leave-one-out
- We considered **different values for k** ($k=1,3,5,9,101$)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>NER</th>
<th>TPR</th>
<th>TNR</th>
<th>PREC\text{POS}</th>
<th>PREC\text{NEG}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic 9-NN</td>
<td>90.1%</td>
<td>89.5%</td>
<td>90.5%</td>
<td>87.5%</td>
<td>92.1%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[85.7-94.5]</td>
<td>[85.3-93.8]</td>
<td>[86.0-95.0]</td>
<td>[81.6-93.4]</td>
<td>[86.8-97.4]</td>
</tr>
<tr>
<td>Modified 9-NN</td>
<td>91.1%</td>
<td>91.9%</td>
<td>90.5%</td>
<td>87.8%</td>
<td>93.7%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[86.8-95.4]</td>
<td>[87.9-95.9]</td>
<td>[86.0-95.0]</td>
<td>[81.9-93.7]</td>
<td>[89.1-98.4]</td>
</tr>
<tr>
<td>Fuzzy $k$-NN</td>
<td><strong>92.6%</strong></td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
<tr>
<td>LD</td>
<td>89.6%</td>
<td><strong>96.5%</strong></td>
<td>84.5%</td>
<td>82.2%</td>
<td><strong>97.0%</strong></td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[85.0-94.2]</td>
<td>[93.7-99.3]</td>
<td>[79.1-89.9]</td>
<td>[75.2-89.1]</td>
<td>[93.9-100]</td>
</tr>
<tr>
<td>QD</td>
<td><strong>92.6%</strong></td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
<tr>
<td>ANN</td>
<td>91.6%</td>
<td>91.9%</td>
<td><strong>91.3793%</strong></td>
<td><strong>88.8%</strong></td>
<td>93.8%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[87.4-95.8]</td>
<td>[87.9-95.9]</td>
<td>[87.0-95.8]</td>
<td>[84.1-93.4]</td>
<td>[88.2-99.4]</td>
</tr>
</tbody>
</table>
Fuzzy k-NN demonstrated to be robust to k changes, keeping its results invariant

<table>
<thead>
<tr>
<th></th>
<th>TRUE LABELS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td>ESTIMATED</td>
<td>82</td>
</tr>
<tr>
<td>LABELS</td>
<td>4</td>
</tr>
</tbody>
</table>

**Indexes | Average Index | Confidence Interval (CI = 95%)**
---|----------------|------------------|
Accuracy | 92.6%          | [88.5-96.7]      |
Sensitivity | 95.3%          | [91.8-98.9]      |
Specificity | 90.5%          | [86.0-95.0]      |
\( \text{PREC}_{POS} \) | 88.2%          | [82.3-94.1]      |
\( \text{PREC}_{NEG} \) | 96.3%          | [93.2-99.4]      |
Results: Confusion Matrix

- Fuzzy k-NN demonstrated to be **robust** to k changes, keeping its results invariant

<table>
<thead>
<tr>
<th>CONFUSION MATRIX</th>
<th>TRUE LABELS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td>ESTIMATED LABELS</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indexes</th>
<th>Average Index</th>
<th>Confidence Interval (CI = 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>92.6%</td>
<td>[88.5-96.7]</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>95.3%</td>
<td>[91.8-98.9]</td>
</tr>
<tr>
<td>Specificity</td>
<td>90.5%</td>
<td>[86.0-95.0]</td>
</tr>
<tr>
<td>PREC_{POS}</td>
<td>88.2%</td>
<td>[82.3-94.1]</td>
</tr>
<tr>
<td>PREC_{NEG}</td>
<td>96.3%</td>
<td>[93.2-99.4]</td>
</tr>
</tbody>
</table>
Fuzzy k-NN demonstrated to be robust to k changes, keeping its results invariant.

<table>
<thead>
<tr>
<th>CONFUSION MATRIX</th>
<th>TRUE LABELS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td>ESTIMATED LABELS</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indexes</th>
<th>Average Index</th>
<th>Confidence Interval (CI = 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>92.6%</td>
<td>[88.5-96.7]</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>95.3%</td>
<td>[91.8-98.9]</td>
</tr>
<tr>
<td>Specificity</td>
<td>90.5%</td>
<td>[86.0-95.0]</td>
</tr>
<tr>
<td>PREC\textsubscript{POS}</td>
<td>88.2%</td>
<td>[82.3-94.1]</td>
</tr>
<tr>
<td>PREC\textsubscript{NEG}</td>
<td>96.3%</td>
<td>[93.2-99.4]</td>
</tr>
</tbody>
</table>
Results: Confusion Matrix

- Fuzzy k-NN demonstrated to be robust to k changes, keeping its results invariant

<table>
<thead>
<tr>
<th>CONFUSION MATRIX</th>
<th>TRUE LABELS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td>ESTIMATED LABELS</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indexes</th>
<th>Average Index</th>
<th>Confidence Interval (CI = 95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>92.6%</td>
<td>[88.5-96.7]</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>95.3%</td>
<td>[91.8-98.9]</td>
</tr>
<tr>
<td>Specificity</td>
<td>90.5%</td>
<td>[86.0-95.0]</td>
</tr>
<tr>
<td>PREC_{POS}</td>
<td>88.2%</td>
<td>[82.3-94.1]</td>
</tr>
<tr>
<td>PREC_{NEG}</td>
<td>96.3%</td>
<td>[93.2-99.4]</td>
</tr>
</tbody>
</table>
Performing a **student t-test** between all pairs of classifiers, no relevant differences emerged

- All implemented classifiers result comparable for the considered problem

The robustness showed by Fuzzy k-NN to k changes is not verified in the classic and the modified k-NN, that leads to different results according to the value of k

Moreover the output given by Fuzzy k-NN can be used to investigate the relationship between these values and lung cancer stages
Classifiers Comparison

- Performing a **student t-test** between all pairs of classifiers, no relevant differences emerged
  - All implemented classifiers result comparable for the considered problem
- The **robustness** showed by Fuzzy k-NN to $k$ changes is not verified in the classic and the modified $k$-NN, that leads to different results according to the value of $k$
- Moreover the output given by Fuzzy $k$-NN can be used to investigate the relationship between these values and lung cancer stages
Classifiers Comparison

- Performing a **student t-test** between all pairs of classifiers, no relevant differences emerged
  - All implemented classifiers result comparable for the considered problem
- The **robustness** showed by Fuzzy k-NN to $k$ changes is not verified in the classic and the modified $k$-NN, that leads to different results according to the value of $k$
- Moreover the output given by Fuzzy $k$-NN can be used to investigate the **relationship between these values and lung cancer stages**
Comparison to current diagnostic methods

- The use of an electronic nose as lung cancer diagnostic tool is reasonable if it gives some advantages compared to current diagnostic techniques.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PREC&lt;sub&gt;POS&lt;/sub&gt;</th>
<th>PREC&lt;sub&gt;NEG&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>Nd</td>
<td>75%</td>
<td>66%</td>
<td>Nd</td>
<td>Nd</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td></td>
<td>[60-90]</td>
<td>[55-77]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET</td>
<td>Nd</td>
<td>91%</td>
<td>86%</td>
<td>Nd</td>
<td>Nd</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td></td>
<td>[81-100]</td>
<td>[78- 94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-Nose</td>
<td>92.6%</td>
<td>95.3%</td>
<td>90.5%</td>
<td>88.2%</td>
<td>96.3%</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>[88.5-96.7]</td>
<td>[91.8-98.9]</td>
<td>[86.0-95.0]</td>
<td>[82.3-94.1]</td>
<td>[93.2-99.4]</td>
</tr>
</tbody>
</table>

- Electronic nose results better in terms of performance.
- Electronic noses are cheaper, smaller (and thus eventually portable), very fast and non invasive instruments.
The use of an electronic nose as lung cancer diagnostic tool is reasonable if it gives some advantages compared to current diagnostic techniques.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PREC\textsubscript{POS}</th>
<th>PREC\textsubscript{NEG}</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT Confidence Interval</td>
<td>Nd</td>
<td>75% [60-90]</td>
<td>66% [55-77]</td>
<td>Nd</td>
<td>Nd</td>
</tr>
<tr>
<td>PET Confidence Interval</td>
<td>Nd</td>
<td>91% [81-100]</td>
<td>86% [78- 94]</td>
<td>Nd</td>
<td>Nd</td>
</tr>
<tr>
<td>E-Nose Confidence Interval</td>
<td>92.6% [88.5-96.7]</td>
<td>95.3% [91.8-98.9]</td>
<td>90.5% [86.0-95.0]</td>
<td>88.2% [82.3-94.1]</td>
<td>96.3% [93.2-99.4]</td>
</tr>
</tbody>
</table>

- Electronic nose results better in terms of performance
- Electronic noses are cheaper, smaller (and thus eventually portable), very fast and non invasive instruments
Comparison to current diagnostic methods

- The use of an electronic nose as lung cancer diagnostic tool is reasonable if it gives some advantages compared to current diagnostic techniques.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PREC&lt;sub&gt;POS&lt;/sub&gt;</th>
<th>PREC&lt;sub&gt;NEG&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>Nd</td>
<td>75% [60-90]</td>
<td>66% [55-77]</td>
<td>Nd</td>
<td>Nd</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PET</td>
<td>Nd</td>
<td>91% [81-100]</td>
<td>86% [78-94]</td>
<td>Nd</td>
<td>Nd</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-Nose</td>
<td>92.6% [88.5-96.7]</td>
<td>95.3% [91.8-98.9]</td>
<td>90.5% [86.0-95.0]</td>
<td>88.2% [82.3-94.1]</td>
<td>96.3% [93.2-99.4]</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Electronic nose results better in terms of **performance**.
- Electronic noses are **cheaper**, **smaller** (and thus eventually portable), **very fast** and **non invasive** instruments.
Lung Cancer Identification by an Electronic Nose based on an Array of MOS Sensors

Extension and further directions of research

1. Improvement of Sensors Technology:
   - Development of longer-lyfe and more stable sensors
   - Development of hybrid systems, able to provide both selective and sensitive abilities

2. Improvement of Olfactory Signal Analysis techniques and Classification Algorithms

3. Exploration of Informations hidden in the Olfactory Signal
   - Analysis of the olfactory patterns' changes due to surgery
   - Variation of VOCs in the breath before and after the surgery
     - It could turn out to be useful for therapy
   - Individuation of risk factors connected to lung cancer
   - Involving a larger population, partitioning it according to different stages and using the Fuzzy output information, it would be possible to study the possibility of early diagnose
Extension and further directions of research

1. Improvement of Sensors Technology:
   - Development of longer-lyfe and more stable sensors
   - Development of hybrid systems, able to provide both selective and sensitive abilities

2. Improvement of Olfactory Signal Analysis techniques and Classification Algorithms

3. Exploration of Informations hidden in the Olfactory Signal
   - Analysis of the olfactory patterns' changes due to surgery
   - Variation of VOCs in the breath before and after the surgery
   - It could turn out to be useful for therapy
   - Individuation of risk factors connected to lung cancer
   - Involving a larger population, partitioning it according to different stages and using the Fuzzy output information, it would be possible to study the possibility of early diagnose
Extension and further directions of research

1. Improvement of **Sensors Technology**:
   - Development of longer-lyfe and more stable sensors
   - Development of hybrid systems, able to provide both selective and sensitive abilities

2. Improvement of **Olfactory Signal Analysis techniques and Classification Algorithms**

3. Exploration of **Informations hidden in the Olfactory Signal**
   - Analysis of the olfactory patterns' changes due to surgery
     - Variation of VOCs in the breath before and after the surgery
     - It could turn out to be useful for therapy
   - Individuation of **risk factors** connected to lung cancer
   - Involving a larger population, partitioning it according to different stages and using the Fuzzy output information, it would be possible to study the possibility of **early diagnose**
Extension and further directions of research

1. Improvement of Sensors Technology:
   - Development of longer-lyfe and more stable sensors
   - Development of hybrid systems, able to provide both selective and sensitive abilities

2. Improvement of Olfactory Signal Analysis techniques and Classification Algorithms

3. Exploration of Informations hidden in the Olfactory Signal
   - Analysis of the olfactory patterns' changes due to surgery
   - Variation of VOCs in the breath before and after the surgery
   - It could turn out to be useful for therapy
   - Individuation of risk factors connected to lung cancer
   - Involving a larger population, partitioning it according to different stages and using the Fuzzy output information, it would be possible to study the possibility of early diagnose
Extension and further directions of research

1. Improvement of Sensors Technology:
   - Development of longer-lyfe and more stable sensors
   - Development of hybrid systems, able to provide both selective and sensitive abilities

2. Improvement of Olfactory Signal Analysis techniques and Classification Algorithms

3. Exploration of Informations hidden in the Olfactory Signal
   - Analysis of the olfactory patterns' changes due to surgery
   - Variation of VOCs in the breath before and after the surgery
   - It could turn out to be useful for therapy
   - Individuation of risk factors connected to lung cancer
   - Involving a larger population, partitioning it according to different stages and using the Fuzzy output information, it would be possible to study the possibility of early diagnose
1. Improvement of Sensors Technology:
   • Development of longer-lyfe and more stable sensors
   • Development of hybrid systems, able to provide both selective and sensitive abilities

2. Improvement of Olfactory Signal Analysis techniques and Classification Algorithms

3. Exploration of Informations hidden in the Olfactory Signal
   • Analysis of the olfactory patterns' changes due to surgery
     • Variation of VOCs in the breath before and after the surgery
     • It could turn out to be useful for therapy
   • Individuation of risk factors connected to lung cancer
   • Involving a larger population, partitioning it according to different stages and using the Fuzzy output information, it would be possible to study the possibility of early diagnose
Extension and further directions of research

1. Improvement of Sensors Technology:
   - Development of longer-lyfe and more stable sensors
   - Development of hybrid systems, able to provide both selective and sensitive abilities

2. Improvement of Olfactory Signal Analysis techniques and Classification Algorithms

3. Exploration of Informations hidden in the Olfactory Signal
   - Analysis of the olfactory patterns' changes due to surgery
     - Variation of VOCs in the breath before and after the surgery
     - It could turn out to be useful for therapy
   - Individuation of risk factors connected to lung cancer
   - Involving a larger population, partitioning it according to different stages and using the Fuzzy output information, it would be possible to study the possibility of early diagnose
Extension and further directions of research

1. Improvement of Sensors Technology:
   • Development of longer-lyfe and more stable sensors
   • Development of hybrid systems, able to provide both selective and sensitive abilities

2. Improvement of Olfactory Signal Analysis techniques and Classification Algorithms

3. Exploration of Informations hidden in the Olfactory Signal
   • Analysis of the olfactory patterns' changes due to surgery
     • Variation of VOCs in the breath before and after the surgery
     • It could turn out to be useful for therapy
   • Individuation of risk factors connected to lung cancer
   • Involving a larger population, partitioning it according to different stages and using the Fuzzy output information, it would be possible to study the possibility of early diagnose
Thanks!