A Review of Technological Progression from Radiomics to Breathomics for Early Detection of Lung Cancer

Funmilayo S. Moninuola, Oluwadamilola Oshin, Emmanuel Adetiba, Anthony A. Atayero, Ademola Adeyeye, Victoria Oguntosin, Olushola O. James, Anthony A. Adegoke, Obiseye O. Obyemi, Surendra Thakur and Abdultaofeek Abayomi

Abstract: Lung Cancer (LC) is formed by an uncontrolled growth of abnormal cells in the lungs, which alters the body's metabolism. Early detection of LC is very important for a better survival rate and prognosis in patients and persons at risk of the disease. Traditional approaches to LC detection are less sensitive to cancerous cells at the early stage, are invasive and consume a lot of resources in terms of time and cost. Exhaled breath is a good source of Volatile Organic Compound (VOC) biomarkers originating from the airways and tissues. Breath analysis is an ideal approach for the detection of metabolites relating to cancer cells in the lungs. The electronic nose (E-nose) approach for LC detection is apt for detecting VOC biomarkers found in the bronchi. E-nose is non-invasive, inexpensive and easy to use. This review discusses the progression of LC diagnosis from radiomics to breathomics (E-nose). Studies on the potential use of E-nose as an easy and non-invasive tool to analyze VOCs in the exhaled breath of lung cancer patients to characterize suspected lung nodules are also presented. Furthermore, the current gaps in respect of the application of E-nose for early detection of LC are unearthed.

Keywords: Lung Cancer, Early Detection, Radiomics, Breathomics, Electronic Nose

Introduction

Lung Cancer (LC) also referred to as carcinoma of the lung or bronchogenic carcinoma, is an aggressive lung tumor caused by unregulated growth of epithelial cells in the tissues of the lung, usually in the bronchi and the airways. This development may spread to a distant site from the lungs and produce metastatic lesions in the brain, bone, liver or adrenal glands. Often times, lung cancer patients are symptomatic of cough, hemoptysis, dyspnea, chest pain, non-resolving pneumonia, metastatic disease like skeletal pain or neurological issues (Hu et al., 2016). Due to the asymptomatic character of this cancer, most cases are not diagnosed till the illness has advanced significantly and therapeutic treatment is not an option anymore.

The International Agency for Research on Cancer (IARC) and World Health Organization (WHO) estimate that lung cancer is the world’s most prevalent cancer fatality and the 6th leading cause of mortality among all forms of cancer in Africa (Sanni et al., 2018). WHO also predicted that Chronic Obstructive Pulmonary Disease (COPD) will be liable for every third mortality in the globe by 2030. COPD was also proposed as a contributing factor for lung cancer (Welniak et al., 2015). In addition, when localized, only 16% of cases of lung cancer are found, while 22 and 57% are detected at regional and distant stages, respectively. In spite of the understanding that timely diagnosis of lung cancer will have a significant effect on survival rates, there has not been an appropriate diagnostic process that offers proof of early detection as is...
available for other types of cancer such as breast and colorectal (Herbst et al., 2018; Lanni et al., 2018).

Lung carcinoma is traditionally classified histologically into 2 key forms: Small Cell LC (SCLC) and Non-Small Cell LC (NSCLC). Mutations involving EGFR, KRAS and TP53 have also been used to subclassify lung cancers on a molecular level. SCLC, which mostly begins in the wider airways as well as the main and secondary bronchi, accounts for around 10-15% of lung cancers. It is the most severe form of lung cancer and develops faster than NSCLC, frequently metastasizing at the early onset of the disease to other areas of the organ. At the onset of symptoms, most SCLC cases have widespread metastasis. SCLC’s 5-year survival rate (6%) is lower than that of NSCLC (21%) (Linning et al., 2019).

The most prevalent form of lung cancer is NSCLC, grouped on the basis of therapeutic outlook and the cancer cells’ gene expression. Roughly 85% of lung cancer cases fall into this category. Adenocarcinoma (40%), squamous cell carcinoma (25-30%) and large cell carcinoma (10-15%) are the three major NSCLC subtypes.

Many patients diagnosed with NSCLC are at a complicated phase in which it is important to envisage the probability of survival and therapeutic options. Regrettably, during diagnosis, 70% of NSCLC tumors cannot be resected, which is important for a clinical analysis that will disclose molecular details and thus the outcome of the prognosis. Knowing the prognosis helps to assess if seeking a certain procedure is useful (Linning et al., 2019; Vavala and Novello, 2017).

If LC is identified early, at least 50% of the patients will survive and be free of relapse tumors. The significant causes of lung cancer are firsthand smoking of tobacco and secondhand smoking in nonsmokers. Other factors include asbestos contamination, automotive and factory air pollution, as well as arsenic, chromium, nickel, aromatic hydrocarbons and other contaminations (Hecht, 2012).

Recognition of LC in its early stages can be characterized on the grounds of lead time, length of time and biases of tumor selection. If it is discovered at an indolent stage, the survival rate will improve significantly. Chemo and radiotherapy treatments are usually used for SCLC, whereas NSCLC is usually treated with surgery. Research shows that use of imaging techniques such as Chest Radiograph (CRG), Computed Tomography (CT), Low-Dose CT (LDCT), Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) are standard approaches for detecting lung cancer.

Nonetheless, all these procedures are expensive and protracted. Also, because of their invasiveness, they are unresponsive to carcinogenic cells in their initial phase. Biopsy and autoantibodies are other common ways of identifying lung cancer, but they are also expensive, time-consuming and require expert physicians. A non-invasive breath test will have significant promise as means of detecting lung carcinoma (Antoniou et al., 2019). A diagnostic test that can identify lung cancer early before it has spread is effective in minimizing lung cancer death rates. The earlier the detection, the higher the chances of successful treatment.

In this review, we focus on the evolution of existing approaches for early lung cancer detection. The resurgence of interest in early identification of LC and the medical imagery used have raised innumerable issues about this international epidemic. Each of the existing methods for LC detection has particular advantages and weaknesses, hence there is continued focus on developing better approaches. The rest of this review is divided into two sections: The trends in LC diagnosis and detection from customary methods to the most recent improvements in exhaled breath analysis are presented, while the latter section presents the research opportunities and the future direction in this area of research.

**Literature Review**

**Conventional Imaging Approaches**

The traditional methods of identification and characterization for LC include: CRG, CT, LDCT, MRI and PET. The purpose of these techniques is to locate lesions in subjects who are at high risk of LC. While they are common strategies with a wide pool of professional users, they are impaired by demerits, which include a high false rate, radiation exposure, high cost and poor sensitivity for the identification of cancer cells at early stages (Kovalchik et al., 2013; Kumar and Latte, 2019). A key advancement in the traditional method for identifying and characterizing LC is the use of advanced algorithms to interpret clinical images called Radiomics. Radiomics improves the precision of LC nodule detection and enables doctors and radiologists to precisely analyze radiographic scans.

**Chest Radiograph**

A chest radiograph is the traditional screening method for the detection of lesions and pulmonary nodules in patients with LC. It uses X-rays to generate lung images, but is associated with the issues of radiation exposure dose and poor image quality. It has also been reported in the literature that this approach has low sensitivity and is not helpful in reducing LC mortality (Khobragade et al., 2016).

**Computed Tomography**

CT is an imaging screening tool that uses sophisticated X-ray technology to detect the condition of the lungs and provides data on tumor features such as dimensions, classification and growth. In addition to staging of the tumor, the 3D CT image presents examination of the chest wall, diaphragm and...
mediastinum invasion. According to Latifi et al. (2015), LC control is substantially influenced by 4D CT by enabling more accurate targeting of the radiation administered. The key disadvantages of CT remain risky exposure to radiation and the detection of indolent tumors that could lead to over-diagnosis and, in most cases, incur additional cost, anxiety and morbidity for patients (Kovalchik et al., 2013; Ohno et al., 2018; Patz et al., 2014).

**Low-Dose Computed Tomography**

Low-Dose Computed Tomography (LDCT) was introduced for lung imaging in order to overcome the limitations of radiation in CT. LDCT acquires single-slice spiral Computed Tomography (CT) using Multidetector row CT (MDCT) scanners. Studies show that LDCT spots more lung nodules and cancer cells than a chest radiograph. Up to 90% of new LC cases were detected using LDCT. It uses an iterative image approach, which is more suitable for the detection of LC in high-risk subjects. However, LDCT is limited in reducing LC mortality as it produces high false-positive rates and the problem of excessive exposure to radiation still persists (Journy et al., 2015; Marshall et al., 2017; NLSTRT, 2013).

**Positron Emission Tomography/Computed Tomography**

Positron Emission Tomography/Computed Tomography (PET/CT) is a fusion of two oncological visualization procedures, which deliver CT structural data and PET anabolic information to envision cancer cells in 3D or bi-dimensional slices, either individually or fused. F18-Fluorodeoxiglucose (F18-FDG) is the most widely used radiotracer, which enables the identification of tumor cells and glucose-consuming metastases similar to majority of malignant pulmonary lesions (Chicklore et al., 2013; de Guevara Hernández, 2015). PET provides better responsiveness and enhanced precision for LC identification compared to CT. In cases with developed NSCLC for stereotactic radiotherapy, PET plays a significant function in the identification of subjects and specification of treatment sites (Mac Manus et al., 2013). The main restrictions of this method are erroneous false-positive results, imprecise structural details, and time-consuming nature (Montilla-Soler et al., 2018).

**Magnetic Resonance Imaging**

Magnetic Resonance Imaging (MRI) was developed to overcome the drawbacks of PET/CT in the investigation of lung malignancies and other diseases. It is a non-ionization radiation tool that works with the principle of magnetic fields and radio waves. It is interfaced with a computer to produce comprehensive images of the structures within the chest and determine the size, extent and degree of LC spread to adjacent structures (Hochhegger et al., 2011). Due to reduced proton intensity and rapid signal degradation of the lung tissue structure, lung MRI has engineering hurdles. In addition, the quality of lung imaging in MRI depends on the patient's ability to conform to the breath hold instructions (Hochhegger et al., 2011; Biederer et al., 2012).

**Magnetic Induction Tomography**

Magnetic Induction Tomography (MIT) is another type of imagery procedure. MIT is a tomography technique that uses a contactless method for mapping the electrical conductivity of tissue and also imaging the electromagnetic properties of an object using the eddy current effect. It can be used to produce passive electrical 3D images, which include neuroimaging applications, radiology cryosurgery tracking and metal flow modeling in metal working procedures. Intramuscular fat and water-bearing fat-free tissue can be classified by MIT (Han et al., 2016). However, MIT imaging techniques have several issues, including high computation imaging algorithms, hypothetical phantoms of the thoracic cavity, challenging clinical hardware systems and spatial resolution (Xiao et al., 2018).

**Computer Aided Diagnosis (CAD) of Lung Cancer using Radiomics**

Radiomics is the advancement of high-resolution extraction processes and techniques for quantifiable functions that transform images from traditional imagery methods into databases that can be extracted to facilitate clinical decision-making (Gillies et al., 2016). To improve the precision of neoplasm nodule detection, the CAD approach was created to enable doctors and radiologists to precisely analyze radiographic scans. The traditional computer-aided testing and identification systems focused on medical imaging for LC include; CRG, CT, LDCT, MRI and PET. Such techniques are designed to distinguish tumors in subjects that are particularly prone to LC (Örozco et al., 2012). Therefore, before making final decisions, radiotherapists utilize the device data to ascertain disease state, since the machine is deemed useful for the identification of possible nodules. The key phases used in radiomics are segmentation, feature extraction, classification and nodule detection. Initial radiographs of individuals are collected from the LC testing database of radiographs with illustrated marked-up lesions such as the Lung Image Database Consortium (LIDC) (Demir and Yılmaz Çamurcu, 2015; Junior et al., 2018).

Although radiomics approaches are well-identified methods with a huge network of knowledgeable operators, demerits exist. When analyzing ordinary computed tomography, doctors and radiographers find it tough to distinguish cancerous nodules because the cross-section is complicated. This involves extra efforts in the identification of the LC by radiologists, and hence there is a high risk of error.
Proteomics and Genomics

Proteomics is the systematic analysis of all the sets of proteins present in cells, chromosomes, or tissues. This technique models how the architecture of proteins is modified by disease. The arrangement of proteins varies from a robust cell to a carcinomatous cell as protein varies or changes occur, which may influence the shape and functionality of the cell (Broodman et al., 2017). Proteomics as well as Genomics are used to sync up gene and protein research demonstrated in a specific subset of genes, cells or tissues and to discover potential molecular targets as well. A concurrent assessment of gene and protein tumor molecular expression uses microarray technology. LC related proteins are produced by LC tissues, which in turn appear to develop high-affinity auto-antibodies to these proteins from the immune system.

Blood biomarkers include specific tumor-associated antigen proteins and auto-antibodies that may be observable 1 to 3 years prior to clinical diagnosis (Jain, 2016; Zhong et al., 2006). Prior studies also used algorithmic solutions for binary classifications of LC biomarkers. Carcino Embryonic Antigen (CEA), CYFRA 21-1 (cytokeratin 19 fragment), Neuron-Specific Enolase (NSE), Progastin Releasing Peptide (ProGRP) and Squamous Cell Carcinoma Antigen (SCCA) are some commonly recognized and clinically used LC protein biomarkers, according to Broodman et al. (2017). Other LC protein biomarkers include Antigen-125 (CA-125), Human Epididymis Protein 4 (HE4) and Surfactant Form of Protein B (Pro-SFTPB) (Sin et al., 2013; Taguchi et al., 2013).

Table 1: Some of the CAD using radonics approaches in the literature

<table>
<thead>
<tr>
<th>S/N</th>
<th>Dataset</th>
<th>Segmentation/Feature Extraction Technique</th>
<th>Classifier</th>
<th>Sensitivity (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LIDC</td>
<td>Toboggan Based Growing Automatic TBGA</td>
<td>RNN</td>
<td>97.93</td>
<td>Sankar and George (2021)</td>
</tr>
<tr>
<td>2</td>
<td>LIDC</td>
<td>Lesion segmentation HoG, PCA, Texture</td>
<td>Multiple classifiers voting</td>
<td>100</td>
<td>Saba (2019)</td>
</tr>
<tr>
<td>3</td>
<td>LIDC</td>
<td>K-NN, Naive Bayes, Texture and geometry feature.</td>
<td>SVM and Ada Boost</td>
<td>98.3</td>
<td>Ngi et al. (2019)</td>
</tr>
<tr>
<td>4</td>
<td>DermIS</td>
<td>GLCM, LBP, Color Features</td>
<td>SVM</td>
<td>97</td>
<td>Khan et al. (2019)</td>
</tr>
<tr>
<td>5</td>
<td>LIDC-IDRI</td>
<td>Frangi filter</td>
<td>4-channel CNN</td>
<td>80.06</td>
<td>Jiang et al. (2017)</td>
</tr>
<tr>
<td>6</td>
<td>LIDC</td>
<td>Watershed segmentation</td>
<td>SVM</td>
<td>92</td>
<td>Makaju et al. (2018)</td>
</tr>
<tr>
<td>7</td>
<td>LIDC-IDRI</td>
<td>NIL</td>
<td>Multi crop convolution neural network MC-CNN</td>
<td>77</td>
<td>Shen et al. (2017)</td>
</tr>
<tr>
<td>8</td>
<td>LIDC-IDRI</td>
<td>Watershed</td>
<td>SVM, Naive Bayes and FLD</td>
<td>94.4</td>
<td>Firmino et al. (2016)</td>
</tr>
<tr>
<td>9</td>
<td>LIDC-IDRI</td>
<td>Clustering</td>
<td>Multi view Convolutional Network</td>
<td>90.1</td>
<td>Setio et al. (2016)</td>
</tr>
<tr>
<td>10</td>
<td>LIDC: ELCAP</td>
<td>Discrete wavelet transforms</td>
<td>SVM</td>
<td>90.9</td>
<td>Orozo et al. (2015)</td>
</tr>
<tr>
<td>11</td>
<td>LIDC-IDRI</td>
<td>Rule based</td>
<td>SVM</td>
<td>93.6</td>
<td>Demir and Yilmaz Camurcu (2015)</td>
</tr>
<tr>
<td>12</td>
<td>LIDC-IDRI</td>
<td>Structural</td>
<td>LDA, gradient boosting (GB 10)</td>
<td>82</td>
<td>Jacobs et al. (2016)</td>
</tr>
<tr>
<td>13</td>
<td>Private</td>
<td>Selective filter based (vessel)</td>
<td>Rule based classification (RBC)</td>
<td>90</td>
<td>Jia et al. (2014)</td>
</tr>
<tr>
<td>14</td>
<td>Private</td>
<td>Cylindrical filter and thresholding based</td>
<td>SVM</td>
<td>90</td>
<td>Teramoto et al. (2014)</td>
</tr>
<tr>
<td>15</td>
<td>LIDC</td>
<td>3D block based on</td>
<td>Entropy Analysis</td>
<td>95.28</td>
<td>Choi and Choi (2013)</td>
</tr>
<tr>
<td>16</td>
<td>JSRT</td>
<td>Clustering watershed segmentation</td>
<td>SVM</td>
<td>85</td>
<td>Chen and Suzuki (2012)</td>
</tr>
<tr>
<td>17</td>
<td>LIDC, ANODE9</td>
<td>Stochastic, binarization and active contour model</td>
<td>SVM</td>
<td>89</td>
<td>Keshani et al. (2013)</td>
</tr>
<tr>
<td>18</td>
<td>ELCAP</td>
<td>Frequency domain analysis</td>
<td>SVM</td>
<td>96.15</td>
<td>Orozo et al. (2012)</td>
</tr>
</tbody>
</table>

Some research works based on this approach as reported in the literature are briefly highlighted in Table 1.

Breathomics

Breath analysis, which is also known as breathomics, is a tool that is avant-garde for prompt detection of lung carcinoma. According to Boots et al. (2012), the aim of breathomics research is to identify trends in Volatile Organic Compounds (VOCs) that define pathological metabolic activities in human beings. Breath science began approximately 50 years ago when Pauling used the process of gas chromatography to show that around two hundred and fifty gases make up human breath (Pauling et al., 1971). Nevertheless, it is still in its growing stage, especially with the current advances in the application of artificial intelligence methodologies to analyze the signature of VOC structures in breaths (Boots et al., 2012).
In clinical analysis and environmental investigation, breath scrutiny can be used because it is non-invasive. Application of breath analysis has been explored across different spectrum of diseases. For instance, lung gangrene and necrotizing pneumonia patients’ breath smells like a sewer, and a stench of mute or fishy reek can suggest liver sickness. Furthermore, the breath smells of untreated diabetic patients are often characterized as “rotten apples” because of the prevalence of acetone in their breath, while failing kidneys can be accompanied by a urine-like smell. Thus, it has been established in the literature that analyzing the breath prints of a LC patient offers several potential advantages (as listed below) other than other conventional medical tests (Ahmed et al., 2017; Shehada et al., 2016; Mertin, 2011):

i. For real-time diagnosis and tracking, breath analysis has great potential
ii. In a gas matrix (breath), the calculation of gas-phase analytes is simplified than in more diverse genetic matrices
iii. Expiratory-breath research is uniquely informative, completely non-invasive and an adjunct or precursor to proteogenomics techniques
iv. Unlike other specimens, breath can be collected easily as often as necessary
v. With no unnecessary side effects, it is simple to acquire and painless for the subject

Breath is a mixture of inorganic gas, inert gas and a tiny percentage of the VOC concentration. When breathing, several molecules are emitted into the air in the range of parts per million (ppm) to parts per trillion (ppt) by volume. The VOCs in breath are produced by endogenic or exogenic processes, and their quantitative and qualitative characterization in individuals is not the same.

O’Neill et al. (1988) claimed that 28 “fingerprints”, which include lipid-peroxidation, hexane, aniline methylene and o-toluidine cancer, are present in the respiratory tract of LC victims (Gordon et al., 1985; Sponring et al., 2010). The distinction in the breath of LC and healthier participants is also confirmed by indicating alkanes and benzene compounds among the 22 VOCs for effective identification of LC (Phillips et al., 2003). To date, no less than 3000 VOCs in the breath of healthy volunteers have been reported and a few of the VOCs are Isoprene, Ethane, Pentane, Acetone and etc.

In clinical prognosis, VOCs arise as a result of metabolic reactions and are informative (Phillips et al., 2003; Mochalski et al., 2013; Rocca, 2018; Phillips et al., 1999; Horváth et al., 2005). Because LC cells have specific metabolic features, the chemistry of VOCs in the expiratory breath parallels cell metabolism within the body. These metabolic residual products flow through the bloodstream and are transferred to the lungs, where they are breathed out via the circulation system. This means that alterations in the metabolic processes of the body result in distinctive prints or signatures of VOC. Researchers have also indicated that the unique metabolism of the disease can be found as breath “fingerprints” showing the presence of LC from the study of exhaled breath of LC patients (Sponring et al., 2010; An et al., 2010; Jordan et al., 2010).

Chemical Analysis of Volatile Organic Compound

There are several procedures explained in the literature to sample, investigate and anamnize exhaled VOCs. The Gas Chromatography (GC) technique is the main standard and is mostly used for chemical analysis of VOCs. Proton Transfer Reaction Mass Spectrometry (PTR-MS), Selected Ion Flow Tube Mass Spectrometry (SIFT-MS), Ion Mobility Spectrometry (IMS), Thermal Desorption Gas Chromatography Mass Spectrometry (TD-GC-MS), Solid Phase Microextraction Mass Spectrometry (SPME-MS), Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR-MS) and FT-ICR-MS are other methods used to investigate VOCs in LC.

Breath sampling in the chemical analysis approach is an important stage in the procedure. Initially, the exhaled breath is collected and preserved briefly. Each VOC is evaluated by GC following a desorption process, which is typically followed by mass spectrometry (GC-MS). Depending on the chemical characteristics, the different VOCs are first isolated, sequentially ionized, and classified by their mass-to-charge (m/z) ratio.

The most common sampling method for breath collection is to use Tedlar sampling bags because they are inexpensive, can easily be manipulated, and are reusable. They are produced from inert resources to prevent both diffusion and reactions between the compounds and the bag. Other sampling methods found in the literature include, Mylar Bags, Cold Trap Systems, Bio VOCs and Gas Tight Syringes (Callol-Sanchez et al., 2017; Naklheh et al., 2017; Oguma et al., 2017; Rudnicka et al., 2011). Sampled VOCs are usually processed using different statistical methods. A few of the research papers for the analytical chemistry of VOCs in the respiratory tract are shown in Table 3.

However, these methods of LC detection are not readily available in health care facilities. This is due to the level of sophistication of the equipment required. They are bulky, expensive and only available in specialized laboratories. Also, these procedures require in-depth preparation of the sample, rare skill-sets to interpret the results and are time-consuming. Furthermore, breath VOC analysis at part per billion by volume (ppbv) to part per trillion by volume (pptv) involving GC-MS based methods requires pre-concentration on commencement of the procedure. This can enhance indications of some VOCs and potentially miss other VOCs (Schmidt and Podmore, 2015; Van de Goor et al., 2018). Even though these techniques are effective for early identification of LC, they are not compact, have no flexibility of use, and, as such, cannot be used in hospitals or homes (at the Point of Care (POC)).
Table 2: Some LC detection studies using proteomics and genomics

<table>
<thead>
<tr>
<th>S/N</th>
<th>Biomarker</th>
<th>Tumor class</th>
<th>Percentage of Sensitivity</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DNA methylation</td>
<td>NSCLC adjacent normal tissues</td>
<td>85.2</td>
<td>Lu et al. (2018)</td>
</tr>
<tr>
<td>2</td>
<td>DNA methylation</td>
<td>138 formalin-fixed -116 LC and 22 controls</td>
<td>100</td>
<td>Walter et al. (2018)</td>
</tr>
<tr>
<td>3</td>
<td>miRNA biomarkers</td>
<td>Exosomes from BAL, 4 Tissue samples</td>
<td>N/A</td>
<td>Kim et al. (2018)</td>
</tr>
<tr>
<td>4</td>
<td>Proteomic</td>
<td>13 AC vs. 15 controls</td>
<td>N/A</td>
<td>Carvalho et al. (2017)</td>
</tr>
<tr>
<td>5</td>
<td>Genomic</td>
<td>- 43 SCLC and 96 NSCLC</td>
<td>95</td>
<td>Kim et al. (2014)</td>
</tr>
<tr>
<td>6</td>
<td>Genomic</td>
<td>NSCLC</td>
<td>99.47</td>
<td>Adetiba and Ologbega (2015a)</td>
</tr>
<tr>
<td>7</td>
<td>Proteomic</td>
<td>- Tissue - 13 NSCLC 6 AC and 7 SCC</td>
<td>N/A</td>
<td>Nan et al. (2017)</td>
</tr>
<tr>
<td>8</td>
<td>Proteomic</td>
<td>- BAL</td>
<td>N/A</td>
<td>Hsu et al. (2016)</td>
</tr>
<tr>
<td>9</td>
<td>Liquid biopsy</td>
<td>- 108 patients with malignant nodules and -113 patients with benign lung nodules</td>
<td>75</td>
<td>Shen et al. (2011)</td>
</tr>
<tr>
<td>10</td>
<td>DNA methylation</td>
<td>Plasma and sputum - 150 NSCLC - 60 Control</td>
<td>Spumt - 98 Plasma - 93</td>
<td>Hudbert et al. (2017)</td>
</tr>
<tr>
<td>11</td>
<td>DNA methylation</td>
<td>- 90 suspected LC prospective followed for two years</td>
<td>N/A</td>
<td>Bagheri et al. (2019)</td>
</tr>
<tr>
<td>12</td>
<td>miRNA biomarkers</td>
<td>- Tissue - 14 AC vs. adjacent normal tissue</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>
Electronic Nose Based Volatile Organic Compound Analysis

The E-nose is a system that senses and classifies VOCs in exhaled breath using a sensor array. By using qualitative and pattern-based "breath print" techniques, this technology alleviates the shortcomings of the chemical analysis approach.

The comparison between the human olfactory system and the E-nose is shown in Fig. 1 pictorially. As demonstrated in the top part of Fig. 1, the E-nose consists of a processing unit for sensing, that is, the hardware and the processing unit for automatic pattern recognition. The sensing device consists of an assembly of sensors where, on the basis of chemical properties, each sensor senses a particular biomarker and then transforms it to a digital signal to generate the “breath print”. The signature of each biomarker is then used to create a labeled signature database that would be used to teach and build an automated pattern recognizer. Principal Component Analysis (PCA) is primarily used to obtain main variables while the pattern recognition units employ supervised pattern classification algorithms such as k-Nearest Neighbors (k-NN), Support Vector Machine (SVM) and Artificial Neural Network (ANN) for the classification of datasets (Estakhroyeh et al., 2018). E-nose as a breathomics procedure has phenomenal prospects to promote PoC-centered early detection of lung cancer. This technique is economical, non-invasive, has a quick response time and highly portable. Brief descriptions of some of the current E-nose applications are given in Table 4.

Early diagnosis is supported by E-nose. It is an affordable, fast, portable and non-invasive tool that can detect LC at an early stage at the POC. As shown in Table 5, the practicability of using E-noses for detecting health concerns and detecting diseases in the primary phase has been established by different researchers.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Sensor type</th>
<th>Classifier</th>
<th>Application Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adak et al. (2020)</td>
<td>QCM</td>
<td>ANN-ABC</td>
<td>Alcohols detection</td>
</tr>
<tr>
<td>Moschini et al. (2018)</td>
<td>MOS</td>
<td>PCA</td>
<td>Meat flavor discrimination</td>
</tr>
<tr>
<td>Sarno and Vijaya (2017)</td>
<td>MOS</td>
<td>KNN</td>
<td>Diabetes detection</td>
</tr>
<tr>
<td>Zetola et al. (2017)</td>
<td>QMB</td>
<td>KNN</td>
<td>Tuberculosis detection</td>
</tr>
<tr>
<td>Längkvist et al. (2013)</td>
<td>ZnO</td>
<td>PCA, SVM</td>
<td>Meat spoilage markers detection</td>
</tr>
<tr>
<td>Fedorov et al. (2021)</td>
<td>MQ</td>
<td>LDA</td>
<td>Cooking state of grilled chicken</td>
</tr>
<tr>
<td>Tudu et al. (2009)</td>
<td>MOS</td>
<td>Fuzzy</td>
<td>Quality evaluation of black tea</td>
</tr>
</tbody>
</table>

MOS - Metal Oxide Sensor, QMB-Quartz microbalance, ZnO-Zinc Oxide

<table>
<thead>
<tr>
<th>Author</th>
<th>Sensor type</th>
<th>Classifier</th>
<th>Sample size</th>
<th>Lung Cancer</th>
<th>Control</th>
<th>Model Accuracy (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chang et al. (2018)</td>
<td>MOS</td>
<td>Multilayer Perceptron (MLP)</td>
<td>37</td>
<td>48</td>
<td>75</td>
<td>79</td>
<td>72</td>
<td>- The focus was on iso-butane, ethanol and toluene biomarkers. - Only Caucasian samples. - Chemical noise. - Limited smokers.</td>
<td></td>
</tr>
<tr>
<td>Van de Goor et al. (2018)</td>
<td>MOS</td>
<td>ANN</td>
<td>60</td>
<td>107</td>
<td>83</td>
<td>83</td>
<td>84</td>
<td>- Early-stage tumor participants were not included. - The biomarkers of focus were not mentioned.</td>
<td>- System was trained off device with a shallow learning. - C# based software was used to generate the breath pattern.</td>
</tr>
<tr>
<td>Wong et al. (2018)</td>
<td>MOS</td>
<td>PCA-KNN</td>
<td>33</td>
<td>12</td>
<td>84</td>
<td>Nil</td>
<td>Nil</td>
<td>- System used wild EGFR mutation breath print. - Feature extraction method was not stated. - The model could not train 31.2% of the control. - Researcher encouraged use of larger number of subjects. - Biomarkers was not stated. - Higher sensitivity to stage 1 LC but conflicting sensitivity at higher stadiation. - PCA was used for feature extraction.</td>
<td></td>
</tr>
<tr>
<td>Li et al. (2017)</td>
<td>LDA</td>
<td>Fuzzy</td>
<td>24</td>
<td>13</td>
<td>91.59</td>
<td>91.72</td>
<td>91.58</td>
<td>- Small dataset. - System used wild EGFR mutation breath print. - Feature extraction method was not stated. - The model could not train 31.2% of the control. - Researcher encouraged use of larger number of subjects. - Biomarkers was not stated. - Higher sensitivity to stage 1 LC but conflicting sensitivity at higher stadiation. - PCA was used for feature extraction.</td>
<td></td>
</tr>
<tr>
<td>Shlomi et al. (2017)</td>
<td>Nanomaterial-based sensor array</td>
<td>DFA</td>
<td>16</td>
<td>30</td>
<td>Nil</td>
<td>75</td>
<td>76</td>
<td>- Small dataset. - System used wild EGFR mutation breath print. - Feature extraction method was not stated. - The model could not train 31.2% of the control. - Researcher encouraged use of larger number of subjects. - Biomarkers was not stated. - Higher sensitivity to stage 1 LC but conflicting sensitivity at higher stadiation. - PCA was used for feature extraction.</td>
<td></td>
</tr>
<tr>
<td>Tirusi et al. (2017)</td>
<td>Cyranose 320 sensor</td>
<td>SVM</td>
<td>45</td>
<td>39</td>
<td>87.3</td>
<td>88.9</td>
<td>66.7</td>
<td>- Feature extraction method was not stated. - The model could not train 31.2% of the control. - Researcher encouraged use of larger number of subjects. - Biomarkers was not stated. - Higher sensitivity to stage 1 LC but conflicting sensitivity at higher stadiation. - PCA was used for feature extraction.</td>
<td></td>
</tr>
<tr>
<td>Gasparri et al. (2016)</td>
<td>QMB</td>
<td>PLS-DA</td>
<td>70</td>
<td>76</td>
<td>Nil</td>
<td>58</td>
<td>91</td>
<td>- The biomarkers were not stated. - SiNW FET was fabricated and was used to check breath print of different disease such as asthma.</td>
<td></td>
</tr>
<tr>
<td>Shehada et al. (2016)</td>
<td>Silicon nanowire field effect transistor (SiNW FET)</td>
<td>ANN</td>
<td>149</td>
<td>129</td>
<td>77.1</td>
<td>20</td>
<td>94</td>
<td>- The biomarkers were not stated. - SiNW FET was fabricated and was used to check breath print of different disease such as asthma.</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Summary of some of the E-nose application fields

Table 5: Existing Studies on E-nose based Analysis of Breath VOCs
As shown in Table 5, the majority of the E-noses developed by researchers are dependent on breath samples from Caucasians, which may not correctly identify black people and other races with LC. In addition, the reported E-noses and inferences were trained off-device. The sample sizes utilized in training the existing E-noses were small populations of high-risk patients. Furthermore, shallow machine learning algorithms were largely used to develop the E-noses in the literature, which could possibly hinder the accuracy, sensitivity and specificity that are sufficient for PoC application.

**Research Opportunities and Future Directions**

This literature review shows that the metabolic alterations associated with LC can be captured from breath as VOC biomarkers, with clinical significance for early detection of LC. Chemical analysis and E-nose-based approaches are used to identify and analyse these VOCs. Many improvements have been made on the procedure or sensing material since the inception of this technology to create low-cost, efficient and handheld devices. Nevertheless, the state-of-the-art results on E-nose research have shown that the device is largely application-specific rather than universal. It can detect and discriminate the production profiles of VOC from microbial infections in-situ and other genetic diseases (Phillips *et al*., 1999; Längkvist *et al*., 2013).

Furthermore, it has provided a plethora of benefits to a variety of commercial industries, including the agricultural, biomedical, cosmetics, environmental, food, manufacturing, military, pharmaceutical, regulatory and various scientific research fields. However, further multi-disciplinary research efforts on the technology will culminate in advances that will improve its attributes such as uniformity, consistency, sensitivity, specificity and accuracy. The VOC biomarker method for LC detection offers several advantages over customary imaging techniques by enabling early spotting of malignant nodules at the PoC in
a non-invasive and pain-free manner, with the added benefit of low cost. This advantage makes this approach a more probable LC detection method, especially in third world countries.

Conclusion

This study has shown the resurgence of interests in LC early detection and the trends in the application of radiomics through proteomics/genomics to breathomics (E-nose) for LC detection over the years. However, there have been only a few efforts on the deployment of E-nose for early detection of LC at the PoC. With increasing innovations in sensing and detection technologies, machine learning and signal processing, E-nose device can be used as a PoC tool to provide rapid data processing and real-time results. Thus, if VOC biomarkers are leveraged for early detection of LC, there would be drastic reduction in LC patients’ mortality rate. A non-invasive, highly specific, sensitive, accurate, cheap, portable and easy to operate intelligent E-nose based on breathomics approach has a great potential for improving the medical outlook of LC.

Acknowledgement

This work was carried out at the Advanced Signal Processing and Machine Intelligence Research (ASPMIR) Group Laboratory, Covenant University, Ota, Nigeria. KZN e-Skills CoLab, Durban University of Technology, Durban, South Africa provided support for the publication of this work.

Author’s Contributions

Funmilayo S. Moninuola: Writing of the manuscript
Oluwadamiola Oshin: Contributed to the writing of the manuscript, Critical review and editing of the manuscript.
Emmanuel Adetiba: Designed the research plan and organized the study, contributed to the writing of the manuscript, Editing of the manuscript.
Anthony A. Atayero: Co-designed the research plan and organized the study, Editing of the manuscript.
Ademola Adeyeye: Editing of the medical aspect of the manuscript, critical review of draft article.
Victoria Oguntosin: Contributed to the writing of the manuscript, Critical revision of the article.
Oluwada O. James: Editing of the manuscript and critical review of the final draft.
Anthony A. Adegoke: Critical review of the final draft and galley proof.
Obiseye O. Obiyemi: Critical review of final draft of the article and galley proof.

Surendra Thakur: Critical review and revision of final draft of the article. Source for funding of manuscript publication.
Abdultafoeek Abayomi: Editing of the manuscript and critical review of the final draft.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved

References


Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: images are more than pictures, they are data. Radiology, 278(2), 563-577.


1082


Shehada, N., Cancilla, J. C., Torrecilla, J. S., Pariente, E. S., Brönstrup, G., Christiansen, S., ... & Haick, H. (2016). Silicon nanowire sensors enable diagnosis of patients via exhaled breath. ACS nano, 10(7), 7047-7057.


