Supplementary Information (SI) for Journal of Materials Chemistry B.

Supporting Document

Room Temperature Operated Flexible MWCNTs/Nb_2O_5 Hybrid Breath Sensor for the Non-Invasive Detection of Exhaled Diabetes Biomarker

Gulshan Verma[1], Sonu Sarraf[2], Aviru K. Basu[2], Pranay Ranjan[3], Ankur Gupta[1*]

[1]Department of Mechanical Engineering, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan -342030, India

[2]Quantum Materials and Devices Unit, Institute of Nano Science and Technology, Mohali, Punjab - 140306, India

[3]Department of Material and Metallurgical Engineering, Indian Institute of Technology Jodhpur, Jodhpur, Rajasthan -342030, India

*Corresponding author: ankurgupta@iitj.ac.in

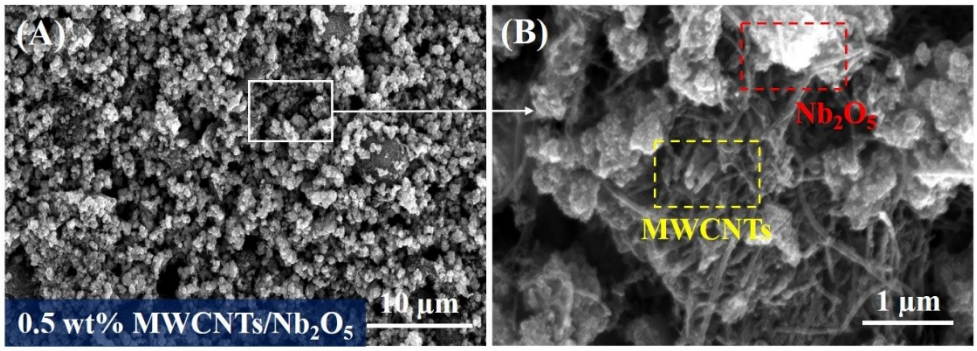

Figure S1. FESEM image of 0.5 wt% MWCNTs/Nb_2O_5.

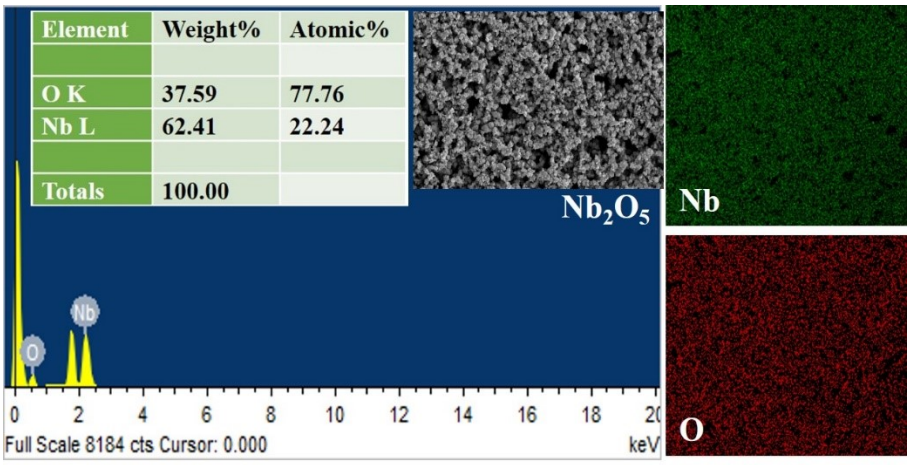

Figure S2. EDS and Elemental Mapping of Nb_2O_5.

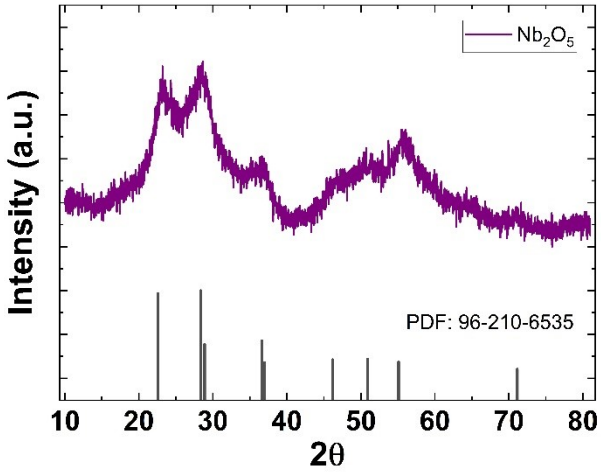

Figure S3. XRD of Nb_2O_5 with PDF card.

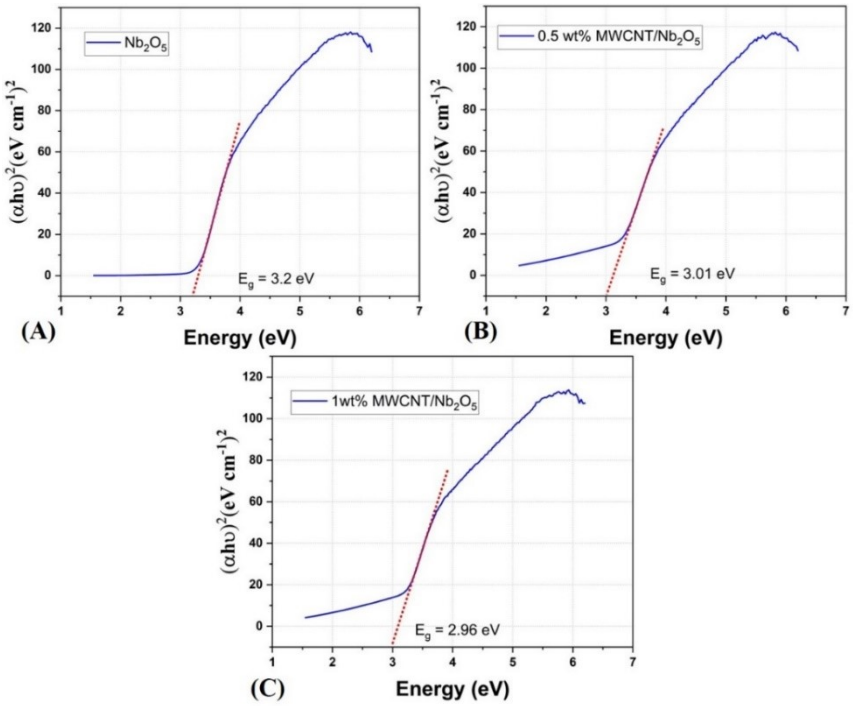

Figure S4. Tauc plot of (A) Nb_2O_5, (B) 0.5 wt% and (C) 1 wt% MWCNTs/Nb_2O_5.

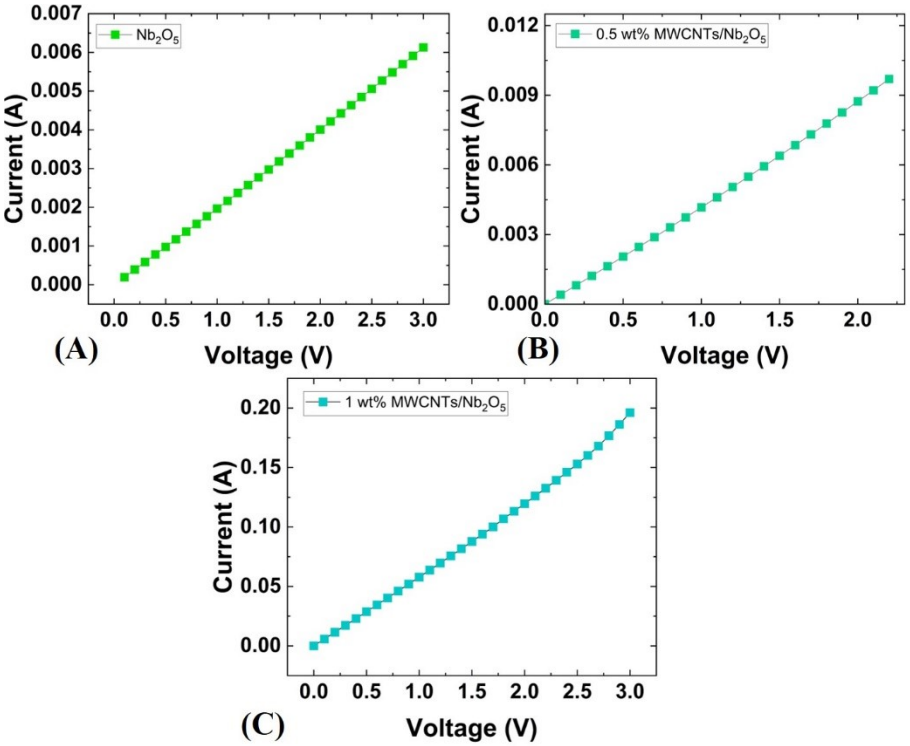

Figure S5. I-V curve of all the three devices.

1 ML Algorithms

1.1 t-Distributed Stochastic Neighbour Embedding (t-SNE)

t-SNE is a technique that is mainly used for dimensionality reduction and data visualization. The unique feature of this technique is it tends to preserve local structure and the global structure simultaneously. The other dimension reduction techniques like PCA or multidimensional scaling, on the other hand, try to preserve the global structure, losing the local structure in the process. Stochastic neighbor embedding (SNE) is a precursor method which forms t-SNE with a t- distributed function. The aim of SNE is to match the distributions (Gaussian distribution is considered) of distances between points in high and low dimensional spaces via conditional probabilities. Consider x_i, the i^{th}, object in high dimension space, and y_i be the the i^{th} object in low dimension space. Then conditional probability of the similarity in high dimensional space, $\mathrm{p}_{\langle \mathrm{j}|\mathrm{i} \rangle}$ is given in equation (2),

$$p_{\langle j|i \rangle} = \frac{\exp(-\left|x_i - x_j\right|^2 / 2\sigma^2)}{\sum_{k \neq i} \exp(-\left|x_i - x_k\right|^2 / 2\sigma^2)} \tag{1}$$

Where σ_i is the variance. This equation (3) conveys that, if two points are relatively close, the above conditional probability is large, but if the points are far apart, $p_{\langle j|i\rangle}$ becomes infinitesimal. Similarly conditional probability in low dimensional space, $q_{\langle j|i\rangle}$ is

$$q_{\langle j|i\rangle} = \frac{\exp(-\left|y_i - y_j\right|^2)}{\sum_{k \neq i} \exp(-\left|y_i - y_k\right|^2)} \tag{2}$$

Variance σ_i, in low dimensional space is 1. These probability functions are matched by using Kullback-Leibler (KL) divergences:

$$C = \sum_i KL\langle P_i \,|\, Q_i \rangle = \sum_i \sum_j \left(p_{\langle j|i\rangle} \log \frac{p_{\langle j|i\rangle}}{q_{\langle j|i\rangle}} \right) \tag{3}$$

This cost function C is sum of KL divergence between P_i the probability distributions of point *i* in high dimensional space, and Q_i probability distributions of *i* with respect to all the other points in low dimensional space. KL divergence is an asymmetric function, which derives small cost, for representing data points located far apart in high dimension as nearby points in low dimension. Perplexity is one more important parameter of t-SNE, that is number of close neighbors each point has. t-SNE uses symmetric version of similarity function, the main feature of symmetric SNE is that $p_{ij} = q_{ji} = 0$ and $p_{ii} = q_{ii} = 0$ for all *i* and *j*. Gradient of cost function is given by,

$$\frac{\partial C}{\partial y_i} = 4 \sum_j (p_{ji} - q_{ji})(y_i - y_j) \tag{4}$$

1.2 k-Nearest-Neighbour classification

k-nearest neighbor rule is a method of supervised classification, which assigns a test sample to the majority category label of its k-nearest training samples. $(x_i, \theta_i)\,; i = 1,2,....n$, where $x_i \in \Re^m$ and θ_i denote the label of x_i for each *i*. Assume that number of classes be C, where $C \geq 2$, i.e each $\theta_i \in \{1,2,....C\}, \forall i$. To assign the label to an incoming point x, following procedure is followed. $d(x, x_i), \forall i = 1,2,....n$, is calculated where d denotes the measured distance. Then these n distances are arranged in ascending order, and the first k distances are considered, where k is a positive integer. Grid search cross validation is used to find the optimal

value of k. If k_i denote the number of points belonging to the i[th] class, then x is classified in class i, if k_i f $k_j, \forall j \neq i$.

1.3 Support vector machines (SVM)

SVM is a supervised machine learning algorithm that can analyze the data for both classification and regression. In a classification problem, the algorithm is employed to deal with categorical data, whereas in regression problem model returns continuous valued output. This algorithm has explicit control over errors and fairly scalable to large dimensional data providing higher accuracies. SVM maximizes the margins between classes by creating hyperplanes. To construct the optimal hyperplane, the following optimization problem is solved.

$$minimize \ \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\xi_i \tag{5}$$

$$\text{Subject to constraints, } y_i(\omega^T x_i + b)^3 \geq 1 - \xi_i, \xi_i \geq 0$$

Where ω^T, $x_i \in \Re^2$ and C is the trade-off parameter between margin and error, ξ_i is slack variable which measure of misclassification error, and y_i is the class label for the i^{th} sample. For support vector regression, the objective function of SVM is generalized by introducing ε-insensitive loss function, which forms ε-tube with optimal hyperplane (best fit line) within it. Hence finding the flattest tube that contains most of training instances with above defined convex ε- insensitive loss function is the optimization problem of SVR.

1.4 Random forest algorithm

Random forest or random decision forest is an ensemble classifier and regressor that makes output predictions by combining outcomes from a sequence of decision trees. Algorithm follows bootstrap sampling of the data sub-sets with random sampling of feature subspaces to build multiple decision trees. Stepwise algorithm of random forest having D training samples and N features is given below:

I. The training data D is randomly sampled into l subsets as $\{D_1, D_2 \ldots D_l\}$ along with replacement, with the help of bagging technique

II. For each training data set D_l, randomly sample feature subspace of m dimension such that $m << N$ at each node and each node is well trained with D_l row sampled data points for m features.

III. For classification decision majority votes among trees are considered whereas for regression mean or median of outputs of each tree are considered.

1.5 Naive Bayes (NB) classifier

Naive Bayes classifier assumes that all features are uncorrelated to each other. For D set of tuples, $D = x_1, x_2, ...x_n$ of n dimensional attribute vector, having divided into m classes $C = C_1, C_2, ...C_m$. Bayesian classifier predicts that X belongs to class C_i assume there are s_l possible values for $x_l, l = 1,2,....n$ and there are K possible values for Y. Naive Bayes' first learns the joint probability distribution $P(X,Y)$ of the input and output by the conditional probability distribution based on the conditional independence assumption:

$$y = argmax\, C_j\, P(Y = C_j) \Pi\, P(X(l) = x(l) \,|\, Y = C_j) \tag{6}$$

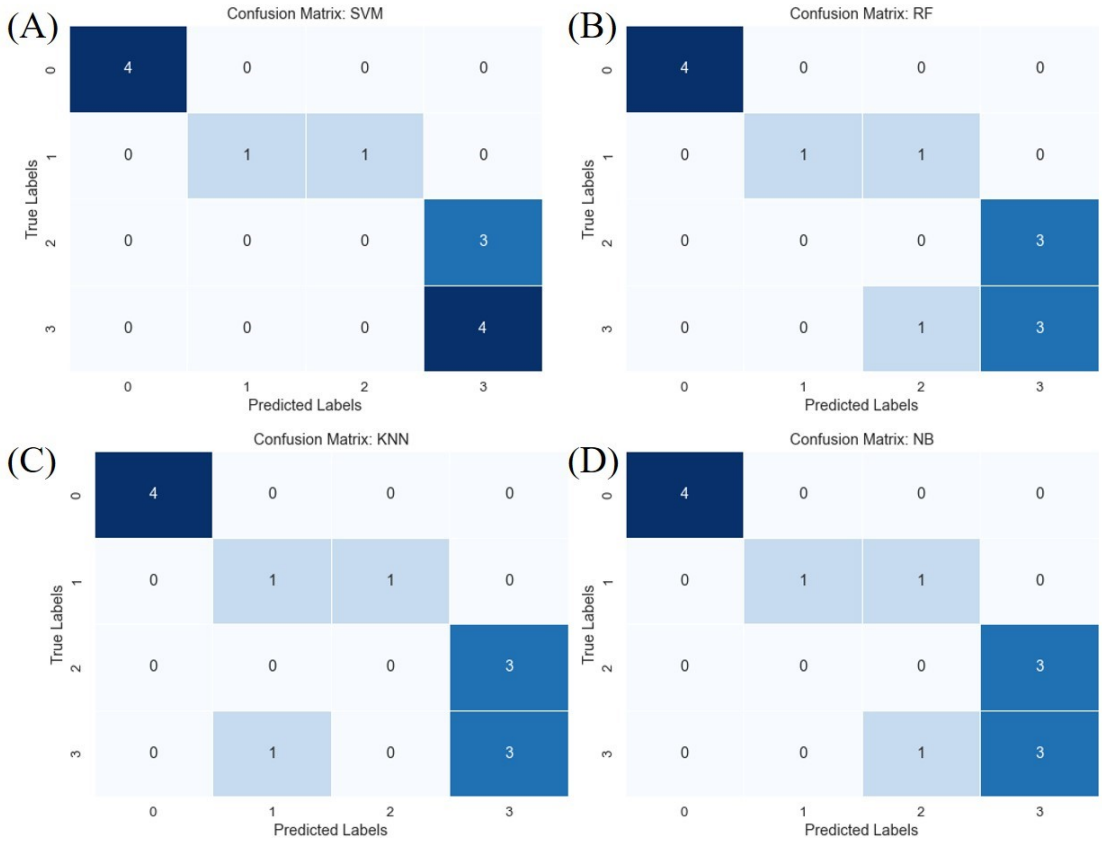

Figure S6. Confusion matrix for SVM, RF, KNN, and NB. (here 0,1,2, and 3 in x and y axis represent acetone, ethanol, methanol and toluene).