Book of Tutorials and Abstracts

European Microbeam Analysis Society

EMAS 2023

17th

EUROPEAN WORKSHOP

on

MODERN DEVELOPMENTS AND APPLICATIONS IN MICROBEAM ANALYSIS

7 to 11 May 2023
at the
Jagiellonian University, Auditorium Maximum
Krakow, Poland

Under the auspices of the Rector of the
Jagiellonian University, Krakow, Poland
Organised in collaboration with the
Institute of Metallurgy and Materials Science of
the Polish Academy of Sciences, Krakow, Poland

EMAS
European Microbeam Analysis Society eV
www.microbeamanalysis.eu/

This volume is published by:
European Microbeam Analysis Society eV (EMAS)
EMAS Secretariat
c/o Eidgenössische Technische Hochschule, Institut für Geochemie und Petrologie
Clausiusstrasse 25
8092 Zürich
Switzerland

ISBN 978 90 8227 6961
NUR code: 972 – Materials Science

MACHINE LEARNING TECHNIQUES IN MICROSCOPIC CHARACTERISATION OF NANO-MATERIALS

Benedykt R. Jany

Jagiellonian University, Marian Smoluchowski Institute of Physics, Faculty of Physics, Astronomy and Applied Computer Science
Lojasiewicza 11, 30348 Krakow, Poland
e-mail: benedykt.jany@uj.edu.pl

Benedykt R. Jany after receiving his PhD in Physics in 2011 from the Jagiellonian University in Krakow in Poland started his work in the field of electron microscopy, nano-technology and surface science. His work was related to the studies of different metal/semiconductor systems (Au/Ge, Au/AIII-BV) at the nano- and atomic scale, using different electron microscopy techniques. As the results of the studies, the new stable form of gold i.e., hexagonal gold (Au hcp) phase was synthesised in the form of 3D nano-structures on germanium surface. This also led to the development of different machine learning based approaches for the characterisation and analysis of the nano-materials.
Currently, after receiving his Habilitation in Physics in 2022 from the Jagiellonian University, he is starting his own research group in the field of Colorimetric Microscopy (C-Microscopy) and Machine Learning.

1. INTRODUCTION

Nowadays variety of different types on nano-materials are used all around us in different fields of science, technology and industry. Because of their scale in single or tens of nanometres the appropriately matched techniques are used for their visualisation and later also characterisation. Microscopy techniques like transmission or scanning electron microscopy (TEM, SEM) together with atomic force microscopy (AFM) are commonly used. Due to the digital acquisition and storage big amount of data are collected by these techniques in the form of microscopy images or hyperspectral data from different type of spectroscopy measurements at nano-scale, etc. It is usually a challenging and tedious task to get reliable information on nano-materials characterisation such as quantitative chemical composition at the nano-scale, structure changes, and various other properties, from all the collected data. In the following, the applications of particular machine learning techniques, which improve and simplify the task of characterisation of nano-materials via selected microscopy techniques will be shown.

2. MACHINE LEARNING SUPPORTED SEM EDS CHEMICAL QUANTIFICATION AT NANO-SCALE

The energy-dispersive X-ray spectrometry in the SEM (SEM-EDS), together with machine learning data processing could be used to quantify chemical composition at the nano-scale [1]. Systems such as AuIn2 metal alloy nano-wires on an InSb substrate (AIIIBV semiconductor) could be successfully quantify. The AuIn2 nano-wires were synthesised in the processes of thermally induces self-assembly of 2 ML (mono-layer) of gold on atomically flat InSb(001) surface at 330 °C in ultra-high vacuum conditions (UHV) [2]. The formed nano-wires are of an average width of ~ 70 nm. The first step towards the quantification is to collect the EDS data at the nano-scale from single nano-wires. The measurement condition (beam current and beam energy) were carefully optimised and verified by CASINO [3] Monte Carlo simulation (see Fig. 1 top). The SEM-EDS data were acquired at 6.5 keV beam energy and 16 nA beam current, using double beam SEM/FIB FEI Quanta 3D FEG microscope equipped with EDAX Ametek Apollo XPP SDD EDS detector with an active area of 10 mm^2. The EDS data were acquired in the hyperspectral mode (EDS Spectrum Image (SI)) i.e., for each x- and y-position a full EDS spectrum was recorded [4].

Figure 1 (bottom) shows collected SEM-EDS results at the nano-scale. The Au-Mα X-ray map shows clearly the nano-wire structures, similar to the simultaneously collected backscattered electron (BSE) image. The EDS spectrum from the AuIn2 nano-wire region (Fig. 1 bottom (2)), shows in addition to the Au and In X-ray signals, also an Sb signal coming from the InSb substrate. This is related to the X-ray generation volume: the EDS signal from the nano-wire region contains also the X-ray signal from below i.e., from the InSb substrate. The EDS signal from the AuIn2 nano-wires is mixed with a signal from the InSb substrate. In the X-ray spectrum there is also a C-Kα signal, which is related to the carbon capping layer and

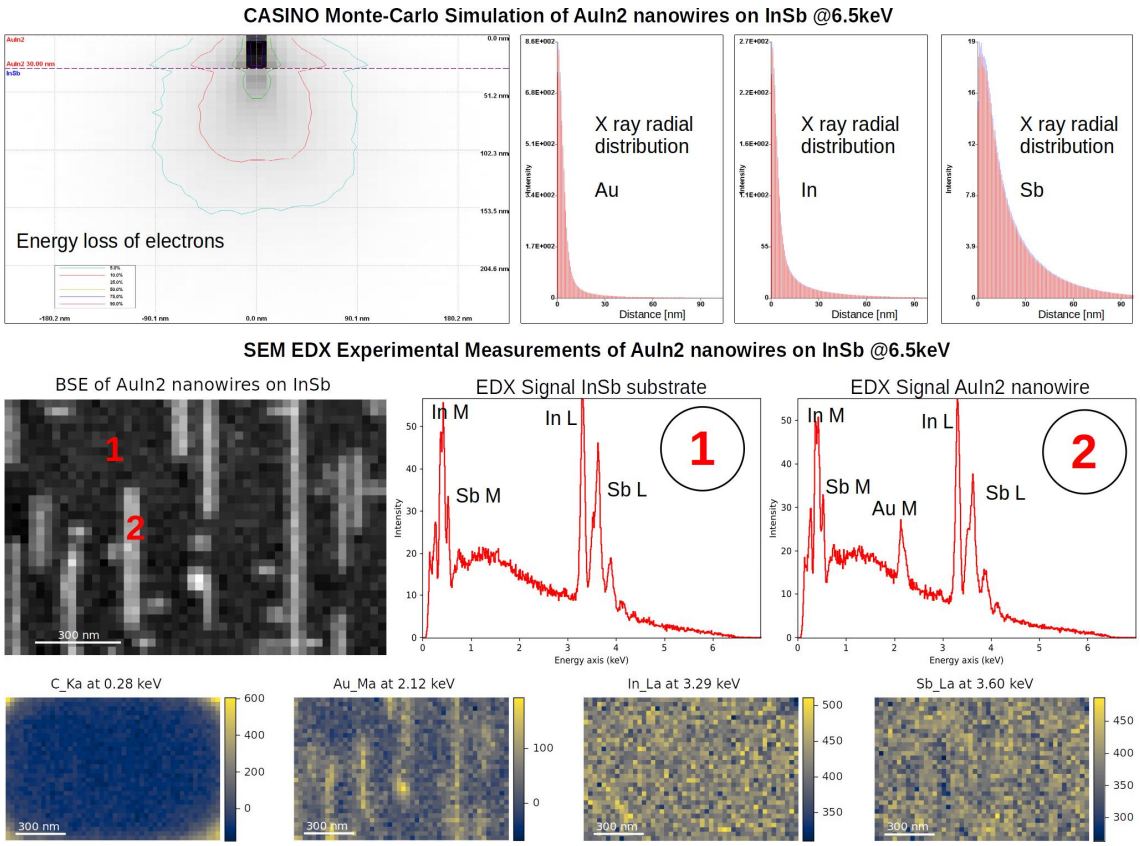

Figure 1. AuIn2 nano-wires on an InSb surface. Top: Results of CASINO Monte Carlo simulation for 6.5 keV electron beam, energy loss of electrons together with X-ray radial distributions for Au, In and Sb. The X-ray signal comes from a depth of around ~ 200 nm. Bottom: Backscattered electron image together with corresponding X-ray maps of C, Au, In and Sb. SEM-EDS spectra are shown for two regions: 1) EDS signal from the InSb substrate region, 2) EDS signal from the AuIn2 nano-wire region. Due to the X-ray generation volume, the EDS signal from the nano-wire region contains also the X-ray signal from below i.e., from the InSb substrate.

contamination. Next, the collected SI EDS data were processed by machine learning using the open software HyperSpy [5]. First, principal component analysis (PCA) was performed to determine the number of components/phases present in the data. Three components were found to have significantly higher variance, as can be seen on the PCA scree plot (Fig. 2). In the next step, non-negative matrix factorisation (NMF) was used to un-mix the data assuming three components found by PCA.

The NMF assumes the data to be non-negative. The results of the NMF decomposition of the SI EDS data is presented in Fig. 2. NMF component maps and corresponding component spectra containing elemental X-ray lines show the following: NMF 1 – InSb substrate, NMF 2 – background (carbon and possibly secondary fluorescence), and NMF 3 – AuIn2 nano-wires. NMF decomposition successfully separated the EDS signal coming from the nano-wires from other EDS signal sources.

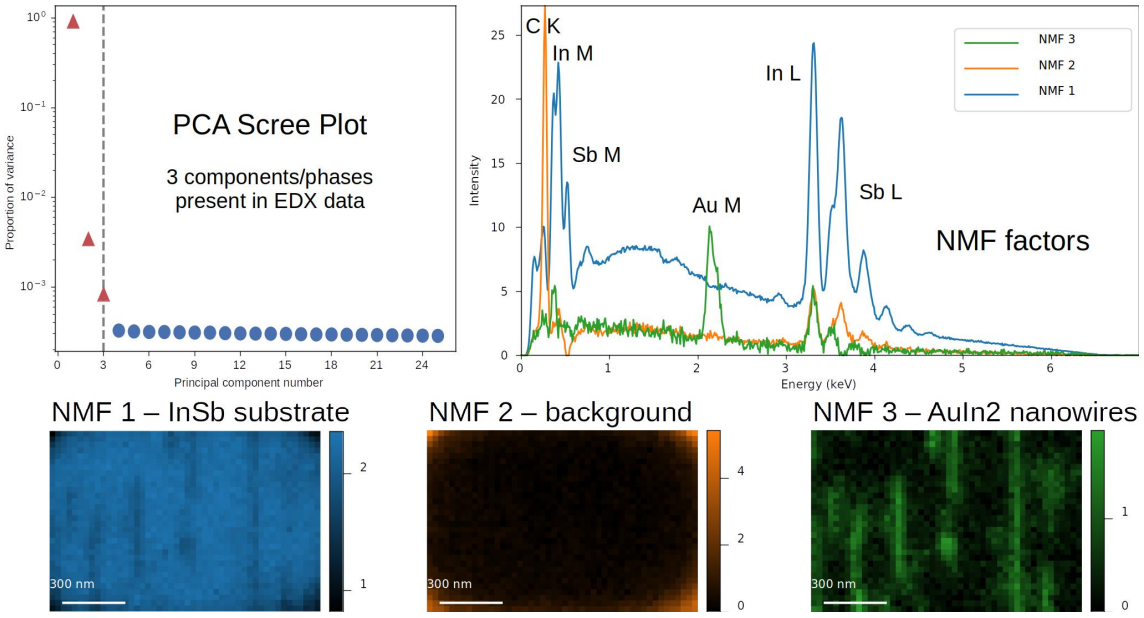

Figure 2. Results of machine learning processing via NMF and PCA of collected SEM-EDS SI data from AuIn2 nano-wires on an InSb substrate. The PCA scree pot shows three components/phases present in the data. Non-negative matrix factorisation (NMF) component maps and corresponding component spectra containing X-ray lines of elements: NMF 1 – InSb substrate, NMF 2 – background and NMF 3 – AuIn2 nano-wires. It shows that NMF separates successfully EDS signal coming from nano-wires.

Subsequently, in order to determine the quantitative chemical composition of the phases extracted by NMF (NMF 1 – InSb substrate, NMF 3 – AuIn2 nano-wires), EDS ZAF standardless quantification was performed using EDAX GENESIS software. The results are presented in Table 1. This shows that the chemical composition determined by NMF decomposition together with ZAF quantification matches the true composition within the estimated uncertainties. The obtained results were further validated by detailed Monte Carlo simulations using DTSA2 software [6] from NIST, as well as additional TEM-EDS cross-sectional measurements [1].

Table 1. Results of EDS ZAF standardless quantification of NMF decomposition spectra.

Decomposition Component	*EDX ZAF Quantification [at. %]*	*True Composition [at. %]*
NMF 1 – InSb substrate	In: 45.0 (4.7) Sb: 55.0 (5.7)	In: 50 Sb: 50
NMF 3 – AuIn2 nanowires	Au: 38.0 (5.9) In: 62.0 (5.9)	Au: 33.33 In: 66.67

The presented machine learning processing approach via NMF of the SEM-EDS SI maps collected at the nano-scale leads to the successful extraction of EDS signal coming from nano-structures, on which later chemical quantification is performed via EDS ZAF method. The spatial limitation of SEM-EDS is greatly improved. The quantitative chemical composition of nano-structures could be successfully obtained by using only simple and achievable SEM-EDS measurements, which could lead to speeding up the measurements and reducing costs in comparison to costly and time-consuming TEM measurements [1]. The presented approach could be used to chemically quantify many different systems at nanoscale by SEM-EDS.

3. MACHINE LEARNING BASED AUTOMATIC MICROSCOPIC IMAGE ANALYSIS

The machine learning approach could be applied to fully automatic microscopic image analysis [7, 8], see Fig. 3. In the presented approach, a moving local window with a defined size is translated in the X- and Y-directions across the microscopic image. The window size (local region of interest on the image) defines the sensitivity scale of the analyses, a 128 pixel size seems to be optimal for images of a size of ~ 2,000 × 2,000 pixels. In the next step, for each local window a 2D power spectrum Fourier-transform is computed. Thus, after this the single image data now consists of two spatial dimensions X and Y and correspondingly two dimensions in reciprocal space KX and KY, this together forms 4D data, very similar to the scanning precession diffraction data (SPED) 4D data. Next, this 4D data is processed by machine learning methods. To determine the components present in the data, the PCA scree plot is calculated, as implemented in HYPERSPY [5] and SCIKIT-LEARN [11]. The programme automatically analyses the PCA scree plot and determines automatically the number of components, which are later used for data decomposition using NMF [5, 11]. This is done fully automatic without any user input. The output of NMF decomposition consists of interpretable spatial maps (loadings) and corresponding 2D power spectra (factors). The image is automatically analysed, i.e., the local features are discovered based of the local FFT, which is sensitive to local orientation and spacing.

This automatic approach was successfully applied to analysis of various microscopic images with and without local periodicity, i.e., SEM, atomically resolved high angle annular dark field (HAADF), scanning transmission electron microscopy (STEM), scanning tunnelling microscopy (STM), and fluorescence microscopy [7, 8]. It could also be successfully applied to other than microscopic images such as photography or a painting, to analyse local features on them [7].

The fully automatic analysis of a single image takes about a minute on an average desktop or notebook computer. This gives the possibility to analyse, within a reasonable time and with computing resources, large amounts of image data taken during measurements on the microscope. The programme outputs a PDF file with a report from analysis. The analysis approach is freely available as a Python analysis notebook and Python programme for batch processing from Mendeley Data [8] (https://doi.org/10.17632/25x46xjyr5.2).

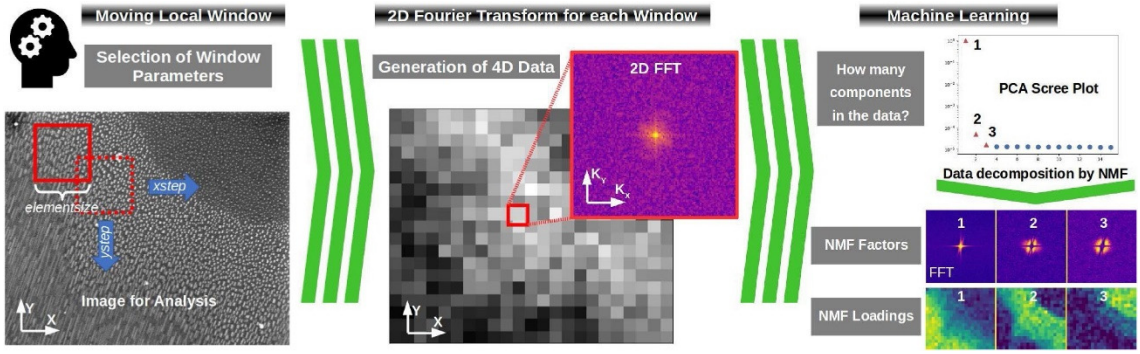

Figure 3. Graphical presentation of the idea of automatic image analysis. A moving local window is applied to the microscopic image and for each local window a 2D power spectrum Fourier-transform is computed resulting in the formation of a 4D data set. Subsequently, the data are fed into machine learning processing: firstly, to determine number of components/fractions in the data, a PCA scree plot is computed; secondly, the data are decomposed blindly into different interpretable components by non-negative matrix factorisation (NMF). The local features in the image are automatically discovered. (From [7] under a Creative Commons license CC by 4.0).

For example, Fig. 4 present the use of the automatic approach to a SEM image (1,662 × 1,370 pixels) of Au-rich nano-structures on a GaSb(001) surface shows three regions with three different types of nano-structures.

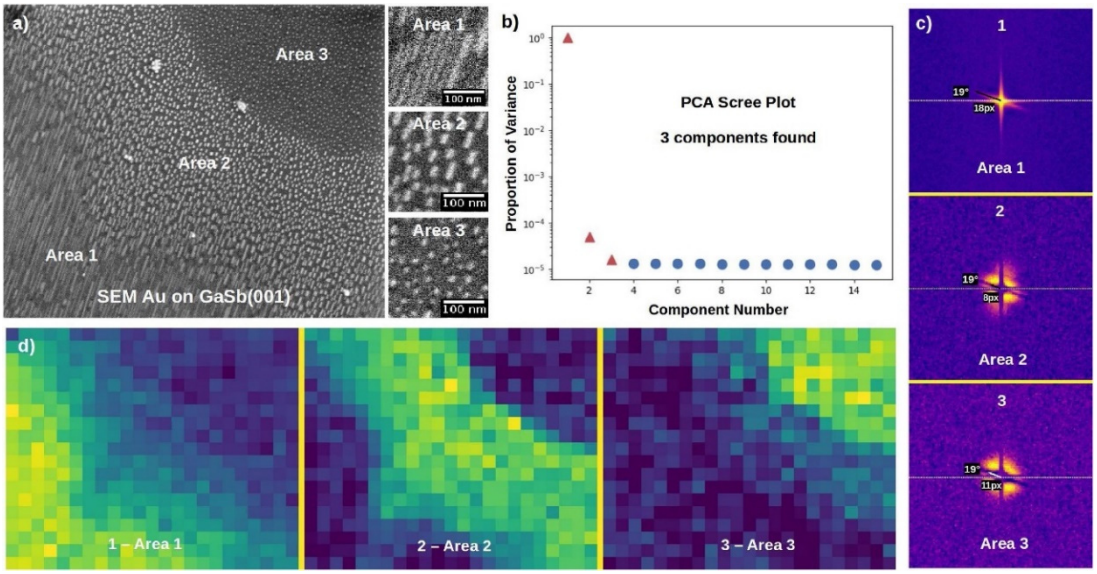

Figure 4. Results of the presented approach applied to a SEM image (1,662 × 1,370 pixels) of Au-rich nano-structures on a GaSb(001) surface. The image shows a) areas of three different morphologies: nano-wires, elongated islands, or symmetric clusters. b) PCA scree plot of the first 15 principal components. The first three components, which exhibit significantly higher variance, are marked with red. c) Non-negative matrix factorisation (NMF) decomposition results of the moving local window Fourier-transform data in the form of decomposition factors (here Fourier transforms). d) Corresponding decomposition loadings (spatial maps). The image width is equal to 3 micrometres. (From [7] under a Creative Commons license CC by 4.0).

These regions are the consequence of different surface stoichiometry related to the non-uniform enrichment of the clean surface in gallium. The analysis, which takes about a minute, automatically finds these three regions in the data (NMF loadings and corresponding NMF factors). From the NMF factors, which are the 2D power spectra characterising each region, one can calculate the characteristic spacing between the nano-structures and the directionality of the structures.

Another example is the application of the automated programme approach to the analysis of an atomically-resolved HAADF STEM image of Au nano-islands grown on a reconstructed Ge(001) surface (Fig. 5). The formed nano-islands are made of two phases of Au, i.e., a cubic fcc-phase and a rare and unusual hexagonal hcp-phase, separated from each other by a defected inter-growth domain along (111) crystallographic planes [9]. The atomically resolved image shows the germanium bulk at the bottom, the nano-island in the middle, and the platinum capping layer on the top. It is not trivial to the naked eye to distinguish in the nano-island the regions of the two different Au phases, i.e., the Au fcc and Au hcp. The analysis, which takes about a minute, automatically finds the regions of the different Au phases. The NMF loadings shows clearly how these regions are located and how they merge into each other. The corresponding NMF factors 2D power spectra characterise each region corresponding to a different phase.

4. ANALYSIS OF CONDUCTIVE AFM CURRENT-VOLTAGE (I-V) DATA BY MACHINE LEARNING

Conductive AFM (C-AFM) is a common technique of local electrical properties characterisation for the nano-scale electronic devices like nano-electrodes. Usually the electrical properties are measured by the conductive AFM tip by acquiring in each image point a full current-voltage (I-V) curve, which forms a hyperspectral data cube. One typical C-AFM measurement in this mode consists of several thousands of I-V curves, which need to be analysed and interpreted. Here, the machine learning methods could be also used to simplify and to enhance the electronic nano-device characterisation [10]. The metallic nano-electrodes grown on an n-doped InP(001) surface were successfully electrically characterised by C-AFM together with machine learning processing [10]. The formed nano-electrodes on InP exhibit two types of morphologies, i.e., “Flat Top” and “Sharp Top” as seen in SEM and AFM topography, see Figs. 6a and 6d. These two different types of morphologies are a consequence of two different types of crystallographic orientation of the metallic nano-electrodes and different atomic structure at the nano-electrode substrate interface, as seen in atomically-resolved HAADF STEM [10]. In consequence, these influences the electronic properties of the nano-electrodes, which were characterised by *in situ* C-AFM measurements in hyperspectral mode, see Figs. 6e and 6h. More than 7,000 I-V curves were collected. The machine learning method k-means clustering, as implemented in SCIKIT-LEARN [11], was used to group the collected I-V data into clusters with similar properties. This is accomplished by iteratively minimising the sum of squares between the points and cluster centres (in the multi-dimensional space) defined as the arithmetic mean of all points,

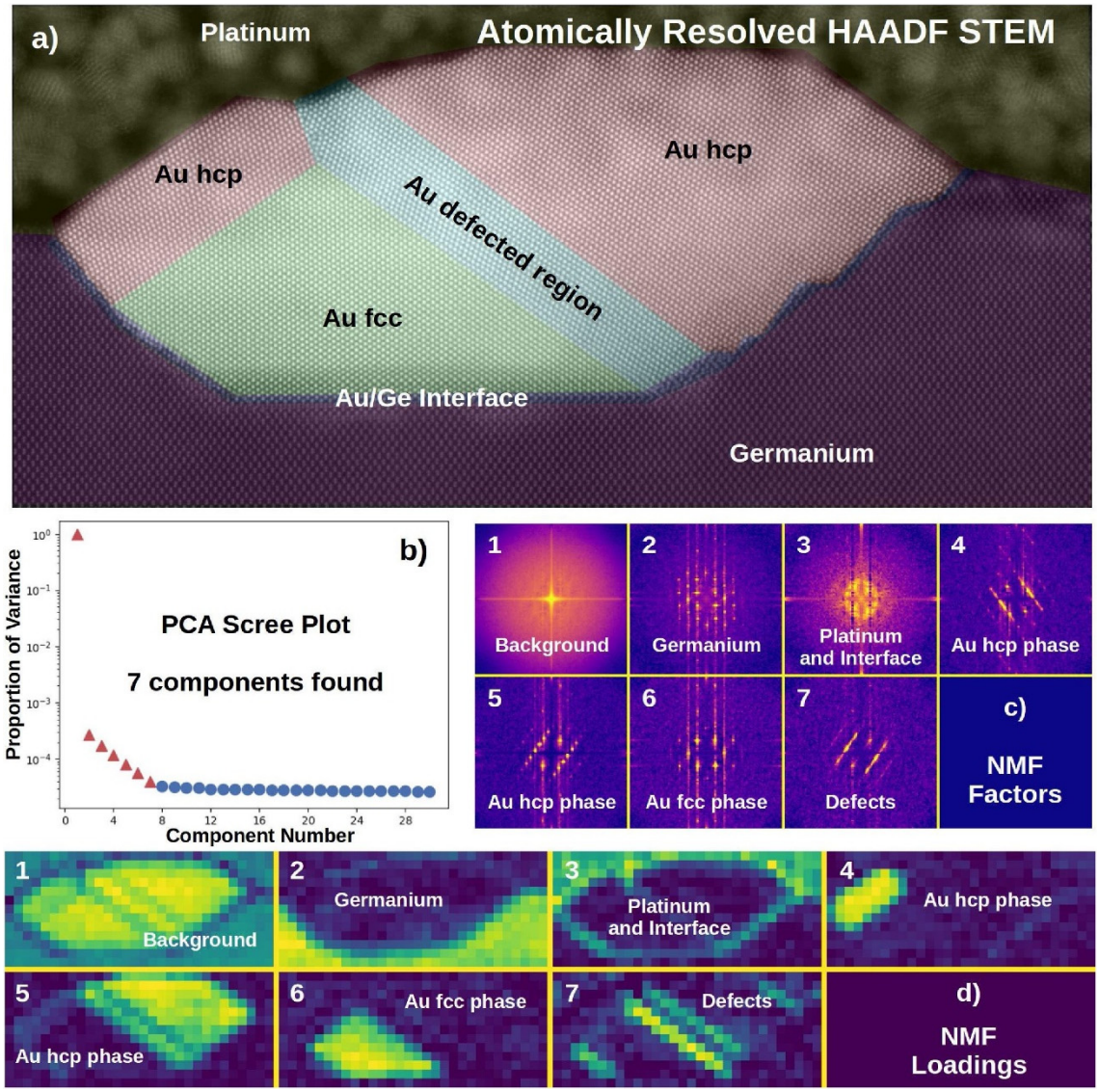

Figure 5. Results of the presented approach applied to an atomically-resolved HAADF STEM image (2,048 × 960 pixels) of a) a nano-island of Au fcc/hcp phases grown on Ge(001) [9] (Creative Commons license by 4.0). b) PCA scree plot of the first 30 principal components. The first seven components exhibit significantly higher variance as compared to the others. c) Non-negative matrix factorisation (NMF) decomposition results of the moving local window Fourier-transform data in the form of decomposition factors (here Fourier transforms). d) Corresponding decomposition loadings (spatial maps). The image width is equal to 49.7 nanometre. (From [7] under a Creative Commons license CC by 4.0).

which belong to the cluster. The C-AFM I-V data were clustered into three groups presented as a map showing different I-V regions, together with corresponding average I-V characteristics in these regions.

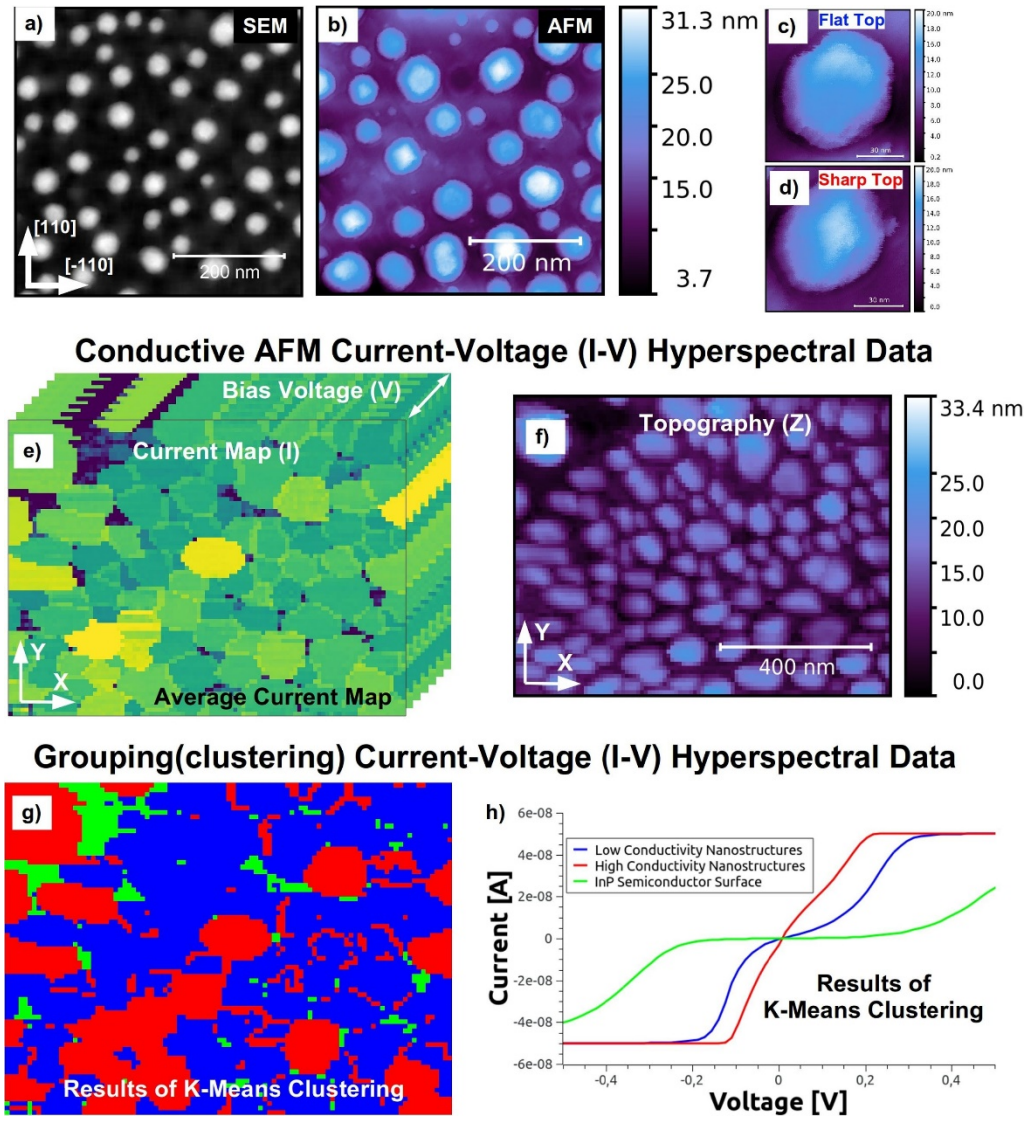

Figure 6. Charactersation of morphology and electrical properties of AuIn2 nano-electrodes grown on n-doped InP(0 0 1) surface. a) SEM, and b) AFM topographies of the nano-electrodes. Two different types of nano-electrodes are visible: c) "Flat Top", and d) "Sharp Top". Results of C-AFM current-voltage (I-V) hyperspectral measurements performed *in situ* after nano-electrode synthesis: e) average current map, and f) corresponding topography. Low conductivity "Flat Top" and high conductivity "Sharp Top" nano-electrodes are visible. Results of grouping (clustering) by machine learning k-means clustering of current-voltage (I-V) hyperspectral data: g) map showing different I-V regions, together with h) corresponding average I-V characteristics in these region. Three different regions are visible: low conductivity nano-electrodes region (blue), which exhibit non-linear I-V behaviour, high conductivity nano-electrodes region (red) with linear I-V characteristic, and an InP surface region (green). It is seen that ~ 70 % of all nano-electrodes are of lower conductivity. (From [10] under a Creative Commons license CC by 4.0).

Three different regions are visible: low conductivity nano-electrodes region (blue), which exhibit non-linear I-V behaviour, high conductivity nano-electrodes region (red) with linear I-V characteristic, and an InP surface region (green). It is seen that ~ 70 % of all nano-electrodes are of lower conductivity. The used machine learning clustering significantly improved the analysis and interpretation of C-AFM I-V measurements [10].

5. *CONCLUSIONS*

The presented selected machine learning based approaches to microscopic data greatly enhance and simplify the task of nano-materials characterisation in the areas of chemical composition quantification [1], structure analysis via microscopic imaging [7], and nano-electrical characterisation [10]. The usage of machine learning techniques to microscopic characterisation of nano-materials allows for reliable extraction of materials properties at the nano-scale level.

The machine learning methods are not limited only to the presented examples, they could be also successfully used to automate and improve different tasks of materials characterisation in the fields of electron backscatter diffraction (EBSD) data analysis or HAADF image quantification [2]. The synergetic coupling between microscopic and machine learning techniques provides new quality and perspectives in materials characterisation.

6. *ACKNOWLEDGMENTS*

This research was supported in part by the Excellence Initiative - Research University Programme at the Jagiellonian University in Krakow.

7. *REFERENCES*

[1] Jany B R, Janas A and Krok F 2017 *Nano Lett.* **17** 6507-7170
[2] Jany B R, Janas A, Piskorz W, Szajna K, Kryshtal A P, Cempura G, Indyka P, Kruk A, Czyrska-Filemonowicz A and Krok F 2020 *Nanoscale* **12** 9067-9081
[3] Drouin D, Couture A R, Joly D, Tastet X, Aimez V and Gauvin R 2007 *Scanning* **29** 92
[4] Jany B R, Janas A and Krok F 2017 *Collected Au/InSb and Au/Ge SEM EDX experimental data at nanoscale as used in retrieving the quantitative chemical information at nanoscale from scanning electron microscope energy dispersive X-ray measurements by machine learning*. http://dx.doi.org/10.13140/RG.2.2.19558.93763
[5] de la Peña F, Ostasevicius T, Fauske V T, Burdet P, Jokubauskas P, Nord M, Sarahan M, Johnstone D N, Prestat E, Taillon J, Caron J, Furnival T, MacArthur K E, Eljarrat A, Mazzucco S, Migunov V, Aarholt T, Walls M, Winkler F, Martineau B, Donval G, Hoglund E R, Zagonel L F, Garmannslund A, Gohlke C and Chang H-W 2017 http://doi.org/10.5281/zenodo.345099

[6] Ritchie N W M 2011 *Microscopy Today* **19** 26-31
[7] Jany B R, Janas A and Krok F 2020 *Micron* **130** 102800
[8] Jany B R 2020 Python Jupyter notebook to perform automatic microscopic image analysis by moving window local Fourier transform and machine learning. Mendeley Data, v2. https://dx.doi.org/10.17632/25x46xjyr5.2
[9] Jany B R, Gauquelin N, Willhammar T, Nikiel M, van den Bos K H W, Janas A, Szajna K, Verbeeck J, Van Aert S, Van Tendeloo G and Krok F 2017 *Sci. Rep.* **7** 42420
[10] Janas A, Piskorz W, Kryshtal A, Cempura G, Belza W, Kruk A, Jany B R and Krok F 2021 *Appl. Surf. Sci.* **570** 150958
[11] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay E 2011 *J. Mach. Learn. Res.* **12** 2825-2830