The general aim of the project is to develop computer programs for more accurate prediction of isoelectric point (*pI*) of protein and peptides. Isoelectric point is the pH at which a particular molecule carries no net electrical charge and it is a critical parameter for many analytical biochemistry and proteomics techniques. Isoelectric point is used to separate protein mixture and identify individual proteins.

Accurate estimation of isoelectric point will be done using state of the art artificial intelligence (AI) technique i.e. deep learning which is implemented in the *TensorFlow* library from Google. It should be stressed that deep learning is a class of artificial intelligence techniques which was used to defeat the masters of chess, Go and recently Texas hold 'em poker, the games in which human performance was out of the reach of AI until now.

Isoelectric point alongside with molecular mass is elementary parameter used for cheap and efficient separation of the protein mixtures. For instance, in each human cell out of ~20,000 proteins the half is expressed – different proteins are present in neurons and different are active in a muscle cell. The number and composition of the proteins in the cell decides about its function and capabilities. Most of the genetic disorders are consequences of a dysfunction of a single protein. In order to design a drug researcher needs to identify, isolate and analyze the protein. Similarly, analysis of most of biological processes requires pure sample of a protein of interest. Therefore, it can be simplified that the first step of many biological analyzes is the isolation of protein from the complex mixture (cell lysate). The most frequently used technique is called two-dimensional gel electrophoresis (2D-PAGE) in which the proteins are separated in two dimension according to the molecular mass and *pI*. In a nutshell, after putting the sample on the gel the electric field is applied and in the result of the proteins spread out on the gel. This process is done in two directions according to the mass and *pI*. At this stage we can specify that given spot contains the proteins e.g. with weight of ~50 kDa and *pI* of 5.5. Having this information we can deduce which proteins can be obtained from a particular area of the gel. This is where the knowledge of theoretical isoelectric point comes handy. Taking into account that the error of current prediction of *pI* from sequence is ± 0.9 we must assume that for our example in the spot on the gel any protein within range of 45-55 kDa and *pI* of 4.6-6.4 can be located. Unfortunately, there can be hundreds such proteins in that range and in order to pinpoint what is in the spot we will need to use more accurate, but also more costly methods (e.g. liquid chromatography–mass spectrometry, LC-MS). If more accurate program for isoelectric point will be developed, than the uncertainty (the error range) of *pI* estimation will be smaller and therefore the whole process of protein isolation will be simpler and shorter on average.

Additionally, there are many other laboratory techniques which will benefit from prior knowledge of isoelectric point. For instance, to limit the complexity frequently LC-MS spectrometry is combined with isoelectric point-based fractionation which allows to improve peptide detection and protein identification. Other case include X-ray crystallography which is used to determine protein structure. This technique relays on forming protein crystals which are built from regularly arranged molecules. Such arrangement allows to amplify the signal and detect the atom positions. Unfortunately, growing the crystals is tedious and very often fails. During this stage the solution conditions (e.g. temperature, pH) and composition of the buffer are manipulated. Analysis of thousands of crystallization trials has shown that in general one should avoid pH similar to isoelectric point of the target protein, therefore more accurate prediction of *pI* from sequence can be invaluable. Moreover, for many diseases the scientists know which of the proteins is responsible for the illness, but without the protein structure they cannot design a proper drug. On the other hand, protein structure determination can take many years of trials costing millions of dollars. Therefore, more accurate prediction of isoelectric point will allow to shorten the process of determining the protein structure.

To sum up, the result of presented research will be a development of new programs for more accurate prediction of isoelectric point. This will simplify the analysis the data coming from a great number of biochemical methods used commonly in practically all laboratories (a simple prototype of the proposed research tool is already online, http://isoelectric.org, and since 2016 it has been used by >100,000 unique users from the whole world). Indirectly this will also accelerate the structure and function determination of the proteins including those which are involved in genetic and neurodegenerative disorders (Alzheimer, Parkinson) and cancer.