Analysis and Prediction of the Spread of COVID-19 in North Macedonia

L.K.LAZAROVA,A.STOJANOVA,N.STOJKOVIKJ, M.MITEVA, M.LJUBENOVSKA

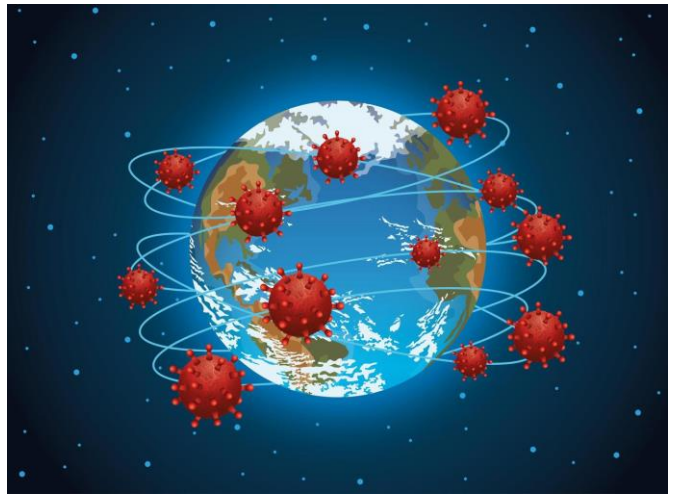
Faculty of Computer Science, Goce Delcev University, Stip, Republic of North Macedonia

Ninth International Conference of FMNS (FMNS-2021)
Modern Trends in Science
15 - 19.09.2021, Blagoevgrad, BULGARIA

INTRODUCTION

During the human history there were many pandemics and epidemics caused by different viruses and all of them had high influence on the people lives.

SARS, EBOLA, MERS and avian influenza, are responsible for the cause of many deaths in many countries worldwide.



The latest pandemic to hit the world, the latest virus that unfortunately humanity has ever heard of, appeared in December 2019. Coronavirus Diseases (COVID-19), a respiratory disease characterized by fever, dry cough, and fatigue, and occasional gastrointestinal symptoms, had its initial outbreak in Wuhan, Hubei Province of China.

There are a lot of victims who did not have the strength to fight against it. As a result of the tragic death cases that occurred because of the COVID-19 virus, The World Health Organization (WHO) announced COVID-19 pandemic on 12 March 2020, when 125.600 confirmed cases were reported from 118 countries and regions from all over the world.

The COVID-19 virus change the life of the whole humanity. It appears almost in the all countries worldwide.

Because of the pandemic the economy system is dealing with a lot of consequences, the governments hand no choice but declaring an emergency state “and had issued regulation with drastic measures for education, business, sports, culture and people.

Many countries, hotels, restaurants, hypermarkets, schools, universities, theatres, churches, and stadiums were closed. There were created new and special rules such as social distancing for communities and people, avoiding meetings in groups, wearing masks and gloves in public places, and isolation at home lockdowns and online working from home became almost normal.

The COVID-19 virus has the most negative impact at the poor countries and the countries with bad public health system.

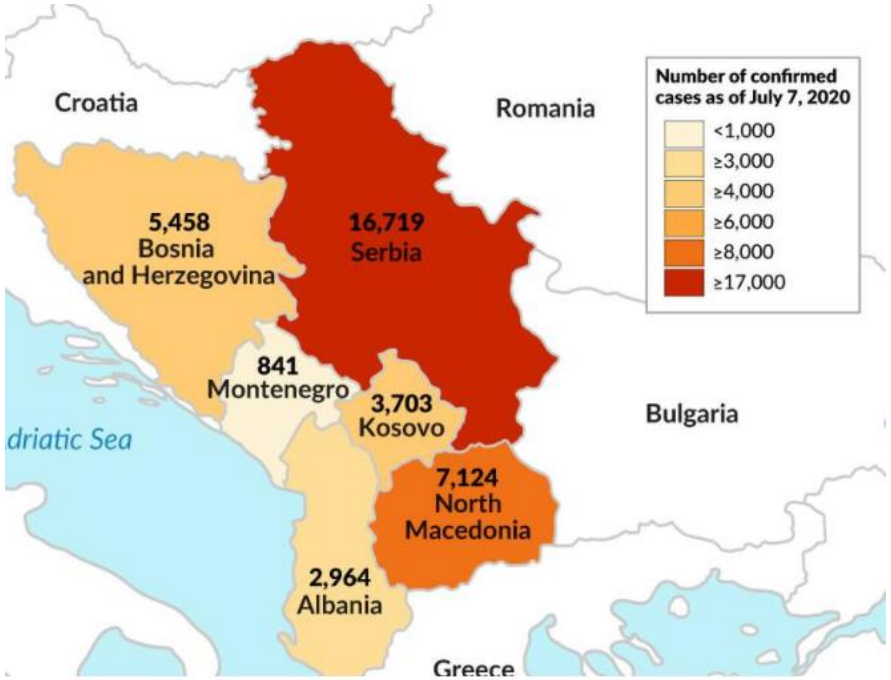
The bad conditions in the hospitals, the lack of medical staff, the very bad treatment of the infected people, no protocol in the process of recovering, not well-planned restrictions and actions for prevention of the infection were the most important reasons for high percent of the infected people and mortality in that countries.

The first case of COVID-19 virus was confirmed in North Macedonia in February 2020. The initial contagion in the country was mainly connected with the COVID-19 pandemic in Italy as there are 70,000 residents of Italy from North Macedonia and resulted in many people returning to North Macedonia, bringing the virus with them.

The stricter preventive measures and sanctions were introduced in March and April 2020.

After that the country went through three waves: October-December 2020, March-April 2021, and the third one started at August 2021 and is still ongoing.

The Republic of North Macedonia, is one of the countries with the highest number of deaths caused by COVID-19 per 1000000 citizens, in the world,



Descriptive statistics of Covid-19 situations in North Macedonia

All descriptive analysis of the Covid-19 situation in North Macedonia is done for four periods.

March-May 2020

No.	Variables	MIN	MAX	MEAN	STD
1	Daily laboratory tests	10	1120	390	206
2	New cases	5	107	30	19
3	Death cases	0	7	2	2
4	Healed cases	0	126	20	25

June-September 2020

No.	Variables	MIN	MAX	MEAN	STD
1	Daily laboratory tests	387	2172	1305	394
2	New cases	11	241	131	40
3	Death cases	0	15	5	3
4	Healed cases	5	534	111	90

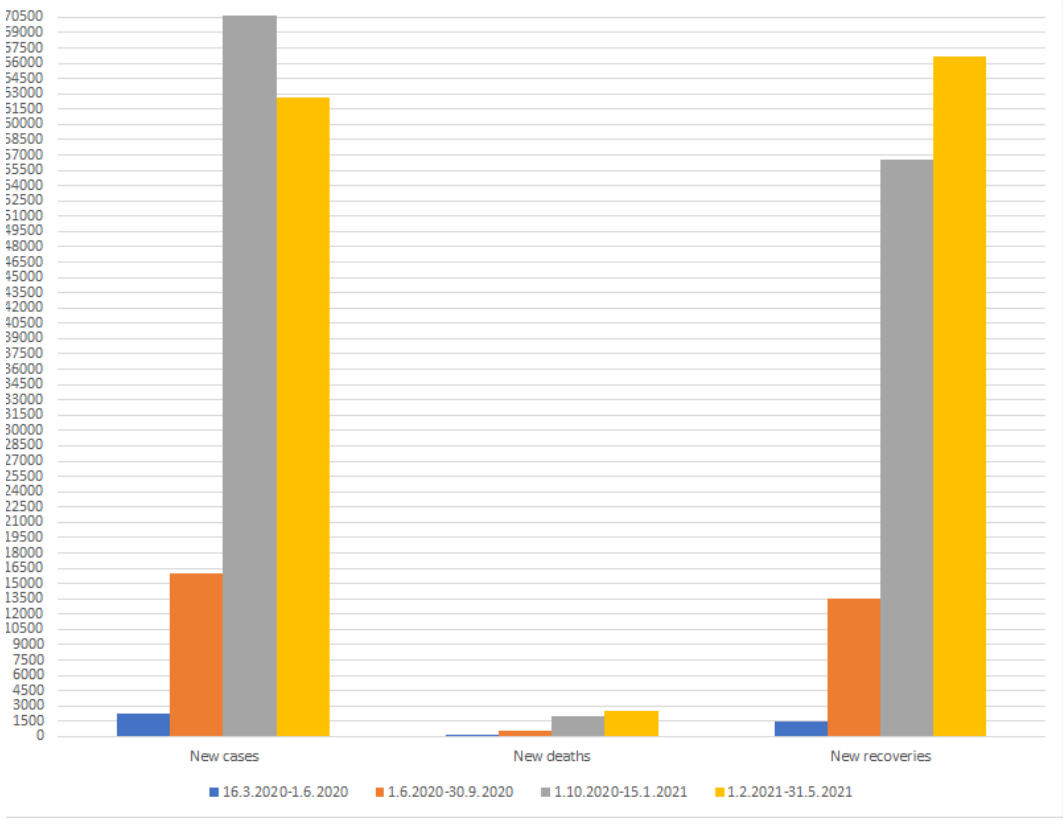
October-December 2020

No.	Variables	MIN	MAX	MEAN	STD
1	Daily laboratory tests	378	3836	2237	828
2	New cases	46	1406	661	370
3	Death cases	3	46	18	10
4	Healed cases	24	2116	529	350

January-May 2021

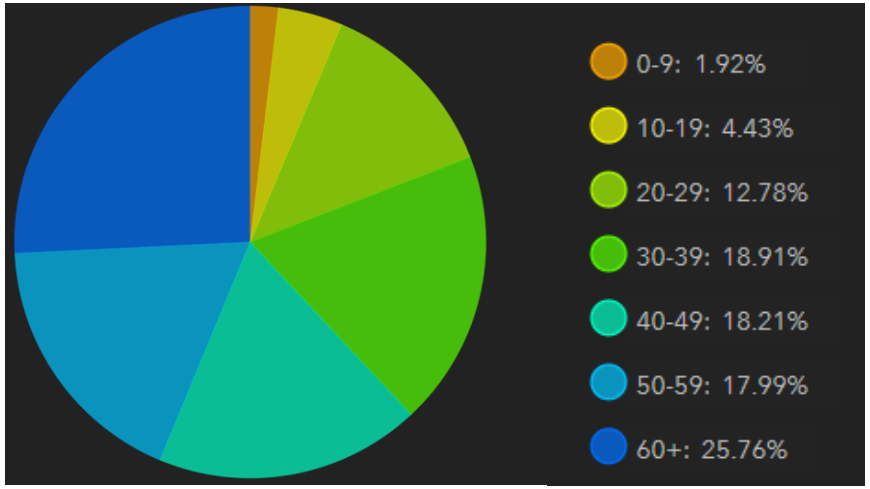
No.	Variables	MIN	MAX	MEAN	STD
1	Daily laboratory tests	572	5569	2852,05	1055,52
2	New cases	3	1511	522,42	426,88
3	Death cases	2	51	21,31	12,11
4	Healed cases	11	3803	555,74	500,52

Descriptive statistics of Covid-19 situations in North Macedonia

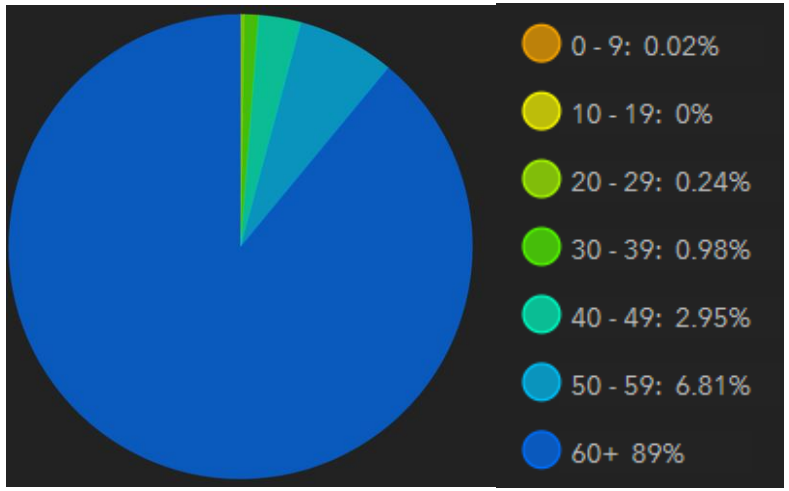


	First Period	Second Period	Third period	Fourth period
New tests	29280	159190	239350	342246
New cases	2214	15935	70725	62690
New deaths	134	603	1953	2557
New recoveries	1521	13509	56582	66689

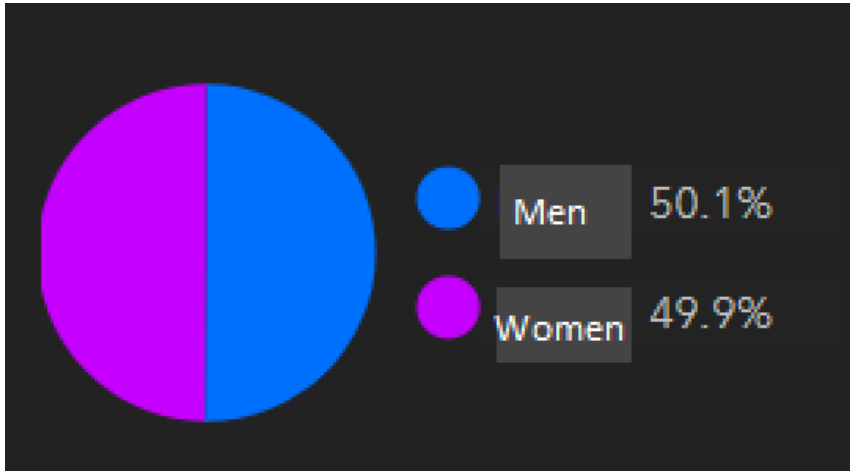
Descriptive statistics of Covid-19 situations in North Macedonia



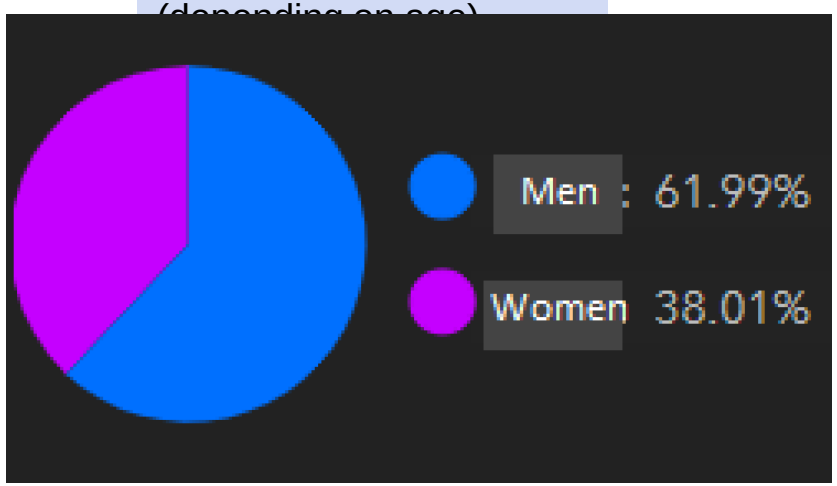
Distribution of infected (depending on age)



Distribution of death (depending on age)

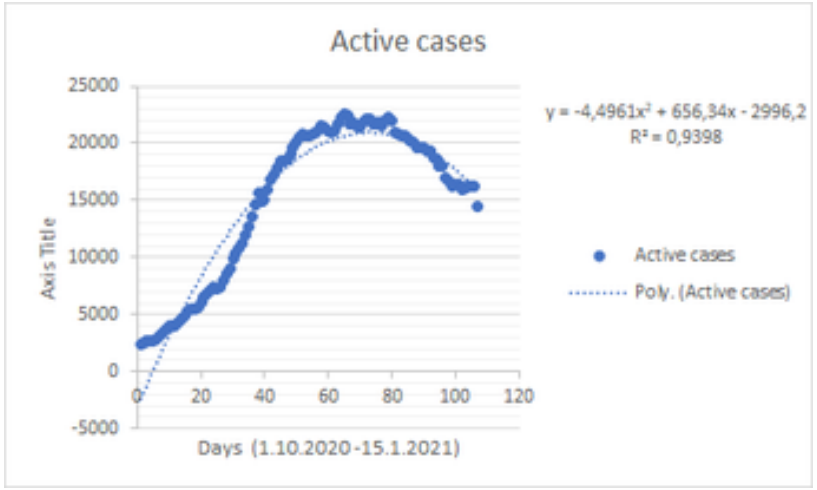


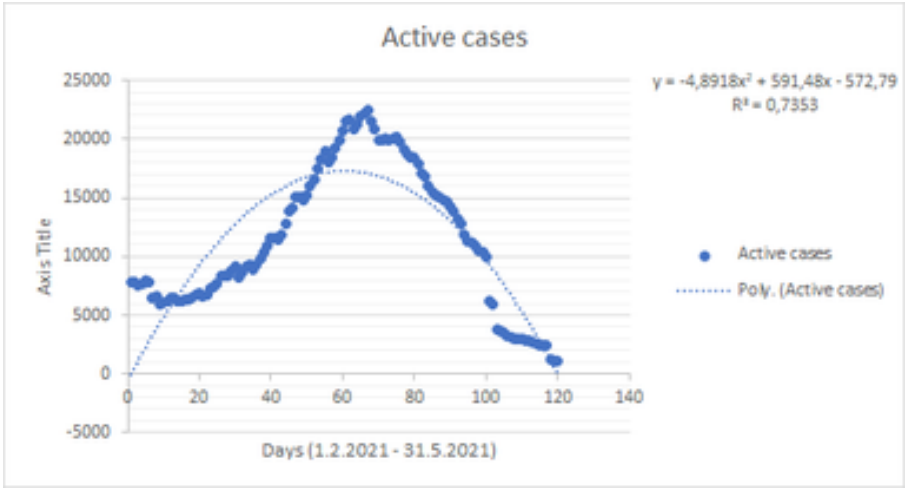
Distribution of infected (depending on gender)



Distribution of death (depending on gender)

Parabolic regression





In order to make predictions about the active cases, the parabolic regression of the active case is done for two waves of COVID-19 virus.
From the coefficient of determination, it can be concluded that the data behave parabolically. So, the number of the active cases can be predicted by the obtain equations.

Machine Learning

For machine learning processing, we used WEKA as a machine learning software. We used 14 different algorithms for classification as a supervised machine learning techniques in order to find the most appropriate one for precise classification and prediction. For classification in WEKA we used 90% of data as training set and 10% for test set.
For precision of classification, we used, Percentage of Correctly Classified Instances and Kappa Statistic. According to values of kappa statistic, value < 0 is indicating no agreement , 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement.

Algorithm used	Percentage of Correctly Classified Instances	Kappa statistic
BayesNet	69.7674 %	0.5847
NaiveBayes	69.7674 %	0.5847
MultilayerPerceptron	74.4186 %	0.5908
SMO	72.093 %	0.5653
KStar	86.0465 %	0.7762
Adaboost M1	74.4186 %	0.5403
Multi-class classifiers	79.0698 %	0.6635
JRip (RIPPER)	65.1163 %	0.3869
OneR	79.0698 %	0.6407
J48	83.7209 %	0.7396
LMT	76.7442 %	0.6315
RandomForest	81.3953 %	0.7085
RandomTree	86.0465 %	0.7737
REPTree	76.7442 %	0.6147

Table 1. New cases of dead

Using data for number of tests per day, new cases of infected, new cases of recovered, total number of infected, total number of recovered and active cases of infected, we try to predict new cases of dead using classification with following algorithms.

According to results obtained, the most precise prediction of 86.0465% derived from algorithm KStar and RandomForest, where kappa statistic is 0.7762 and 0.7737 accordingly.

Machine Learning

In order to predict new cases of recovered, we used number of tests per day, new cases of infected, new cases of dead, total number of infected, total number of dead and active cases of infected.

Algorithm used	Percentage of Correctly Classified Instances	Kappa statistic
BayesNet	60.4651 %	0.5165
NaiveBayes	55.814 %	0.4442
MultilayerPerceptron	67.4419 %	0.5848
SMO	67.4419 %	0.5778
KStar	67.4419 %	0.5825
Adaboost M1	37.2093 %	0.1665
Multi-class classifiers	67.4419 %	0.5825
JRip (RIPPER)	55.814 %	0.405
OneR	60.4651 %	0.4881
J48	72.093 %	0.6404
LMT	60.4651 %	0.4909
RandomForest	65.1163 %	0.5561
RandomTree	60.4651 %	0.5
REPTree	72.093 %	0.6394

Table 2. New cases of recovered

According to results obtained from this table, we can conclude that is difficult to predict new cases of recovered knowing only these data and using these algorithms. The most precise prediction of 72.093 % for Percentage of Correctly Classified Instances has algorithm J48 and REPTree, where kappa statistic is 0.6404 and 0.6394 accordingly.

Machine Learning

In third case we try to predict new cases of infected, we used number of tests per day, new cases of recovered, new cases of dead, total number of recovered, total number of dead and active cases of infected.

Algorithm used	Percentage of Correctly Classified Instances	Kappa statistic
BayesNet	81.3953 %	0.7442
NaiveBayes	81.3953 %	0.7442
MultilayerPerceptron	79.0698 %	0.7077
SMO	86.0465 %	0.8
KStar	79.0698 %	0.7077
Adaboost M1	46.5116 %	0.2764
Multi-class classifiers	86.0465 %	0.8301
JRip (RIPPER)	76.7442 %	0.6702
OneR	79.0698 %	0.6967
J48	81.3953 %	0.738
LMT	86.0465 %	0.8
RandomForest	81.3953 %	0.738
RandomTree	79.0698 %	0.7077
REPTree	79.0698 %	0.7077

According to results obtained and presented in this table, the most precise prediction of 86.0465% derived from algorithm SMO (Support Vector Machine in WEKA), LMT (logistic model trees) and Multi-class classifier, where kappa statistic is 0.8, which can be considered as substantial precision.

Table 3. New cases of infected

Conclusion

In this paper descriptive statistics is made in order to describe the COVID-19 situations in North Macedonia. The analysis and comparations of the situation in different time epidemic periods are done.
From applying the machine learning algorithms used in WEKA for classification, to our collected data, we can conclude that with substantial precision we can classify and thus predict number of dead (with KStar and RandomTree) and number of new cases (with SMO and LMT), but is difficult to predict number of recovered because it probably depends on other parameters.

By using of parabolic regression model and machine learning algorithms, the prediction and directions for future work are given.

So, the analysis can be used as efficient tool to give directions of the authorities to deal with COVID-19 challenges.

Thank You for Your Attention