Значение коэффициента Пуассона полученного композита увеличивается с ростом концентрации вводимого порошка СЭК (таблица 1). Для всех случаев коэффициент Пуансона лежит в пределах значений, характерных для низкоуглеродистых сталей. Плотность также близка к плотностям низколегированных сталей.

Таблица 1 – Значение плотностей и Коэффициента Пуассона для полученных образцов сплавов

Тип матрицы	Тип дисперсионного упрочнителя	Концентрация дисперсионного упрочнителя	Плотность образца, кг/м^3	Коэффициент Пуассона
Fe	WC-TiC-TaC-Cr_3C_2	+0.2 %	7.6117	0.278
Fe	WC-TiC-TaC-Cr_3C_2	+0.5 %	7.6456	0.281
Fe	WC-TiC-TaC-Cr_3C_2	+1.0 %	7.9190	0.287

Работа выполнена при поддержке государственного задания Министерства науки и высшего образования Российской Федерации (проект № FZNS-2024-0013).

УДК 632.952.633

Т.З. Ибрагимов, С.С. Санин, О.М. Рулева,
Л.Г. Корнева, Л.В. Карлова
Всероссийский НИИ фитопатологии
Б. Вяземы, Россия

СИНТЕТИЧЕСКИЕ ДАННЫЕ В ФИТОСАНИТАРНОМ ПРОГНОЗИРОВАНИИ

Аннотация. *Разработка моделей интеллектуального анализа требует наличия больших объемов данных. Синтетические данные – это данные, которые создаются искусственно. Созданы синтетические данные по развитию септориоза листьев пшеницы на основе полевых наблюдений. При увеличении длины ряда точность прогноза развития септориоза увеличилась с 60% (экспериментальные данные, длина ряда 17) до 100% (длина ряда 200 и 400).*

T.Z. Ibragimov, S.S. Sanin, O.M. Ruleva, L.G. Korneva, L.V. Karlova
All-Russia Scientific Research Institute of Phytopathology
B. Vyazemy, Russia

SYNTHETIC DATA IN PHYTOSANITARY FORECASTING

Abstract. *Development of models of intellectual analysis requires the presence of large volumes of data. Synthetic data are data created artificially. Synthetic data on the development of septoria leaf spot of wheat created based on field observations. With an increase in the length of the row, the accuracy of the forecast increased from 60% (experimental data, row length 17) to 100% (row length 200 and 400).*

Синтетические данные – это **данные, которые создаются (генерируются) искусственно.** Они создаются с помощью алгоритмов и используются для широкого спектра действий, в том числе в качестве данных в системах интеллектуального анализа [5,2]. К 2024 году 60% данных, используемых для разработки проектов в области интеллектуального анализа, будут генерироваться синтетическим путем [4].

Методы синтетических данных были впервые предложены Рубином (1993) [8], Литтлом (1993), Райтером (2005). Монография Дрекслера (2011) обобщает некоторые теоретические, практические и политические аспекты разработки и использования синтетических данных [2].

Синтетические данные делятся на два типа в зависимости от того, созданы они на основе реальных наборов данных или нет. Первый тип синтезируется на основе реальных наборов данных. У исследователя имеется набор реальных данных, на основе которых можно построить распределения и структуру этих данных - статистическую модель данных. Синтетические данные генерируются из этой модели и будут иметь аналогичные статистические свойства, что и реальные данные. Второй тип синтетических данных создается посредством моделирования [2].

В наших исследованиях, для создания синтетических данных по развитию септориоза листьев пшеницы, на основе полевых наблюдений, была использована система MOSTLY AI (https://synthetic.mostly.ai) (рис. 1) [4].

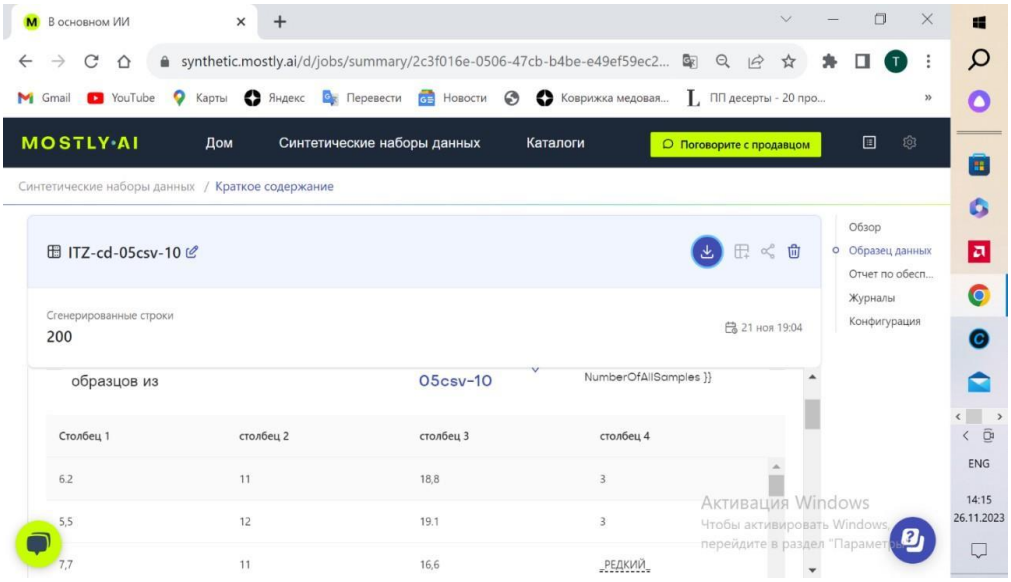

Рис. 1- Интерфейс системы MOSTLY AI

MOSTLY AI – это платформа, которая помогает создавать синтетические данные для использования в машинном обучении, аналитике, тестировании программного обеспечения и т.д. Он генерирует синтетические данные с использованием искусственного интеллекта, который изучает и использует статистические характеристики исходных данных, такие как корреляции, распределения и другие свойства. Это позволяет создавать синтетические данные, которые репрезентативны фактическим данным. В прогнозной аналитике, методы интеллектуального анализа наиболее эффективно используются для классификационного прогнозирования, то есть категориальных данных. Категориальные данные имеют фиксированное количество возможных значений. Чтобы увеличить количество категориальной переменной в наборе данных MOSTLY AI используется специальная функция - разбалансировка данных. Данная процедура может быть применена только к одному столбцу с категориальным типом кодировки. Синтетические данные сохраняют распределение вероятностей категорий и содержат только те категории, которые присутствуют в исходных данных. Редкие категории защищены и обязательно остаются в наборе категориального столбца.

Каждый набор синтетических данных сопровождается отчетом о качестве синтетических данных с диаграммами распределения, корреляции и точности.

С использованием MOSTLY AI были созданы наборы фитосанитарных данных длинной в 100, 200 и 400 наблюдений с разбалансировкой переменной - Развитие болезни в фазе 75.

Точность синтетических данных оценивается путем сравнения распределений статистических характеристик синтетических и исходных данных. В таблице 1 представлена точность синтетических данных.

Таблица 1 - Точность синтетических данных

Переменная	Длина ряда		
	100	200	400
Поражённость листьев в фазе 39	43,8 %	100 %	37,5 %
Средняя температура в фазе 39	43,8 %	31,2 %	43,8 %
Количество дней с осадками в фазе 39	31,2 %	37,5 %	31,2 %
Развитие болезни в фазе 75	68,8 %	100 %	87,5 %
Общая	46,9 %	67,2 %	50,0 %

С помощью многоуровневой нейронной сети прямого распространения (RProp) системы KNIME была оценена их пригодность для прогнозирования развития болезни в фазе 75 для различных наборов данных [3].

При увеличении длины ряда точность прогноза увеличилась с 60% (экспериментальные данные, длина ряда 17 [1]) практически до 100% (длина ряда 200 и 400).

Полученные результаты показывают, что, синтетические данные могут быть использованы для построения фитосанитарных прогнозов в защите растений от болезней.

Список использованных источников

1. Ибрагимов Т.З. Интеллектуальный анализ в фитосанитарии: метод нейронных сетей // Защита и карантин растений. –2022.–№ 10. – С. 8-10.
2. Drechsler J. Synthetic Data Sets for Statistical Disclosure Control, 2011, Springer-Verlag, 981 p
3. Режим доступа: http://www.knime.org.
4. Режим доступа: https://synthetic.mostly.ai.
5. Rubin D.B. Discussion: Statistical Disclosure Limitation // Journal of Official Statistics, 1993, 9 (2), pp.461–468