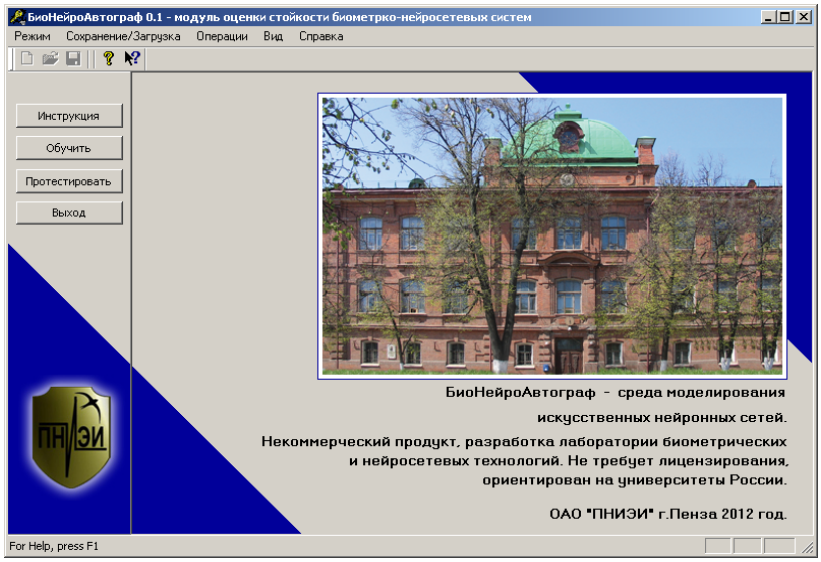
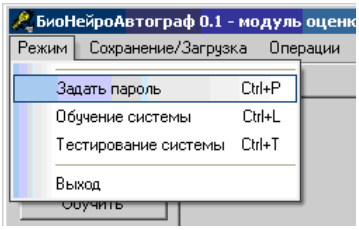
Научно-образовательный центр "Информационная безопасность систем и технологий" ОАО "ПНИЭИ" и ФБГОУ ВПО "Пензенского государственного университета"

Лабораторная работа №6 "Оценка гипотезы нормальности закона распределения расстояний Хэмминга между кодами "Свой" и "Чужой"

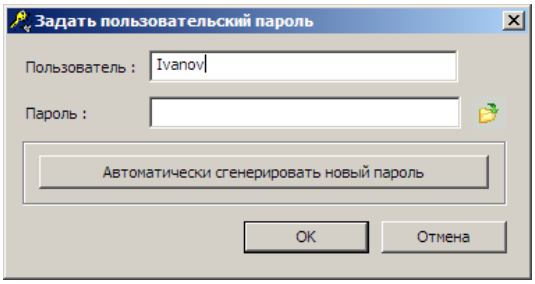
1. Подключите к ПЭВМ графический планшет любой фирмы, установите драйвер графического планшета. Запустите среду моделирования "БиоНейроАвтограф" (файл БиоНейроАвтограф.exe), при этом появится главное диалоговое окно программы.



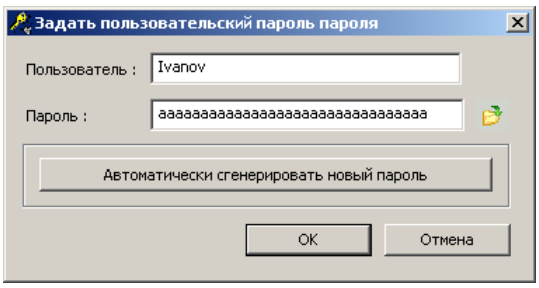
2. Выберите пункт меню "Режим".
3. Выберите режим "Задать пароль".



4. В появившемся диалоговом окне создания пароля в поле "Пользователь" введите свою фамилию или имя, под которым вы будете работать в системе.

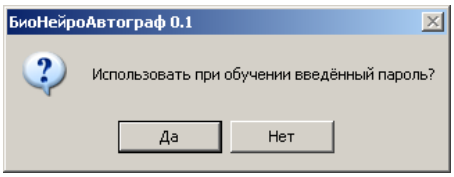


5. Далее в поле "Пароль" задайте пароль из 32-х символов "*аааааа…аааааа*". Пароль вводится в латинской кодировке клавиатуры.

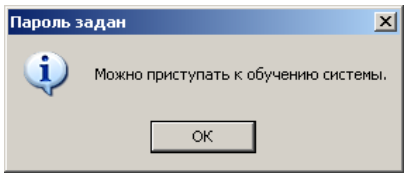


6. Далее нажмите "ОК".

7. В появившемся диалоговом окне нажмите "Да". После этого введённое имя пользователя и пароль будут использоваться при обучении и тестировании системы.



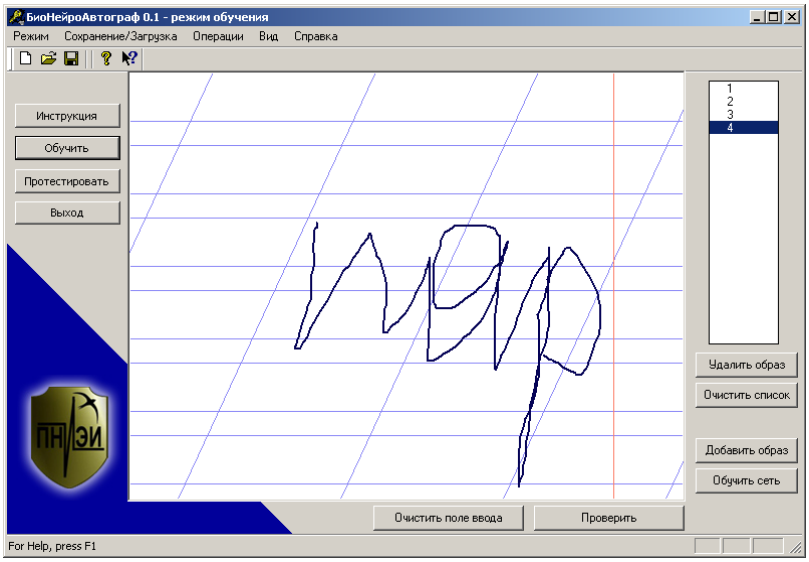
8. Если все пользовательские данные сохранены успешно, то появится сообщение об успешном создании пароля. Нажмите "ОК".



9. После создания пароля можно приступать к обучению системы. Для этого в главном диалоговом окне программы нажмите кнопку "Обучить".

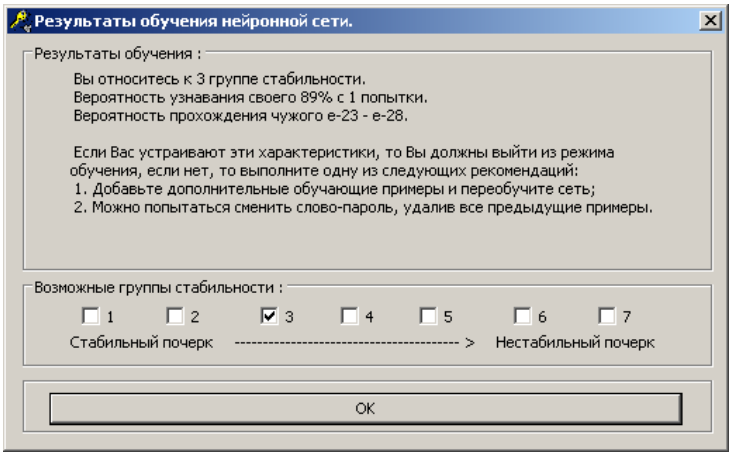
10. Появится диалоговое окно обучения с разлинованным полем ввода рукописных символов/слов. Рукописное слово-пароль вводить с помощью графического планшета.

11. В поле ввода введите один пример рукописного слова-пароля "*пар*" своим почерком, пользуясь манипулятором "мышь" или пером графического планшета (ввод слова печатными буквами не допускается), далее нажмите кнопку "Добавить образ".



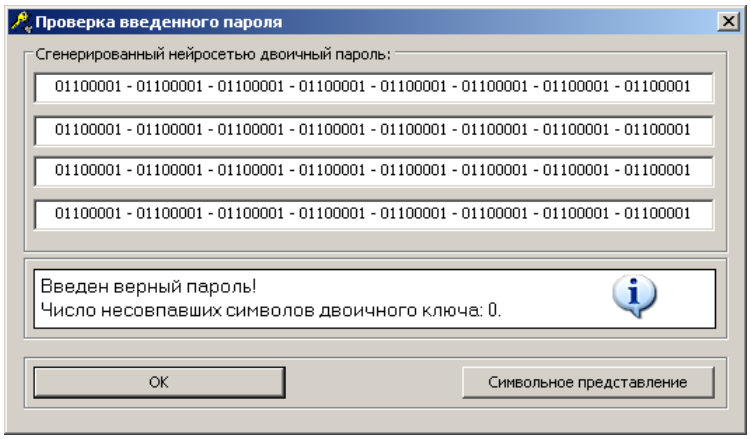
Повторите операцию ввода не менее 8 раз. Рукописные образы нужно писать быстро, опираясь на имеющиеся у вас подсознательные рефлексы, выработанные много лет назад на уроках чистописания.

12. После ввода достаточного количества примеров (8 – 12 рукописных слов) нажмите кнопку "Обучить сеть", при этом начнётся процесс обучения и через несколько секунд появится диалоговое окно с результатами обучения.

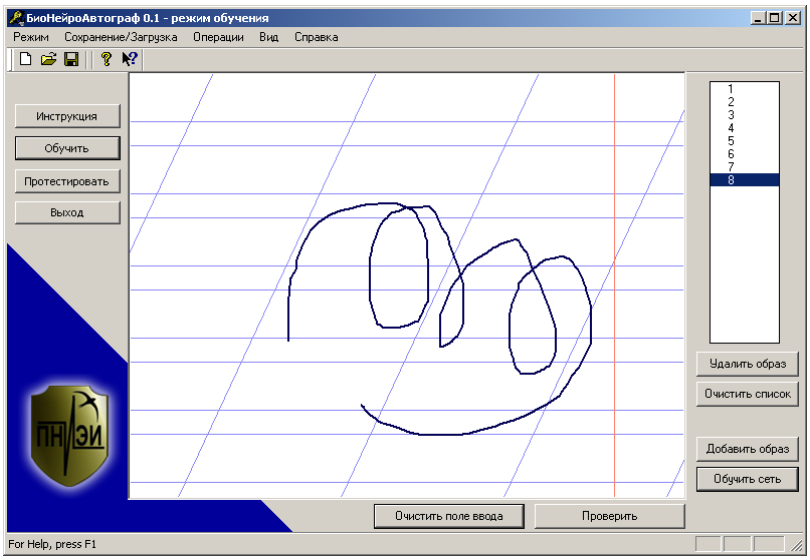


Для закрытия окна нажмите кнопку "ОК"

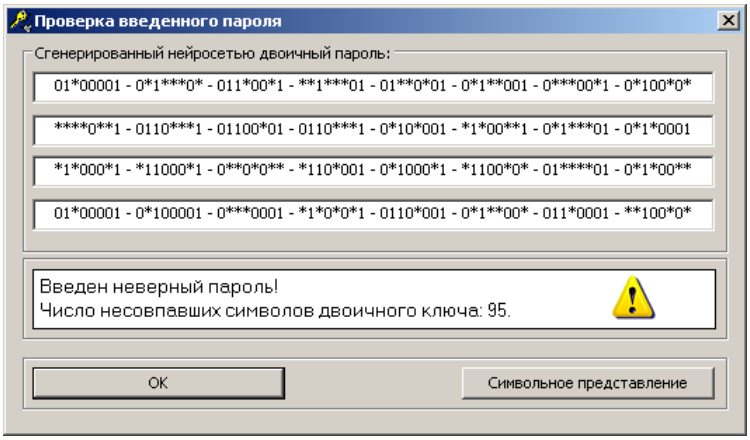
13. Для проверки качества обучения введите контрольный рукописный образ *"пар"* и нажмите кнопку "Проверить". Если средство аутентификации вас узнаёт, то появится сообщение "Введен верный пароль!".



14. Воспроизведите попытки атаки, когда "Чужой" не знает правильный пароль. Для этой цели напишите произвольное слово и нажмите "Проверить".



При этом нейронная сеть перестает узнавать образ. Убедитесь в этом, рассматривая полученный ключ в двоичной и символьной кодировках.

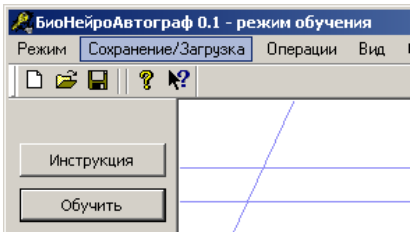


15. Сохраните все обучающие примеры в отдельный файл, выбрав пункт меню "Сохранение/Загрузка" подпункт "Сохранить образы на диск" или нажатием на кнопку с пиктограммой дискеты.

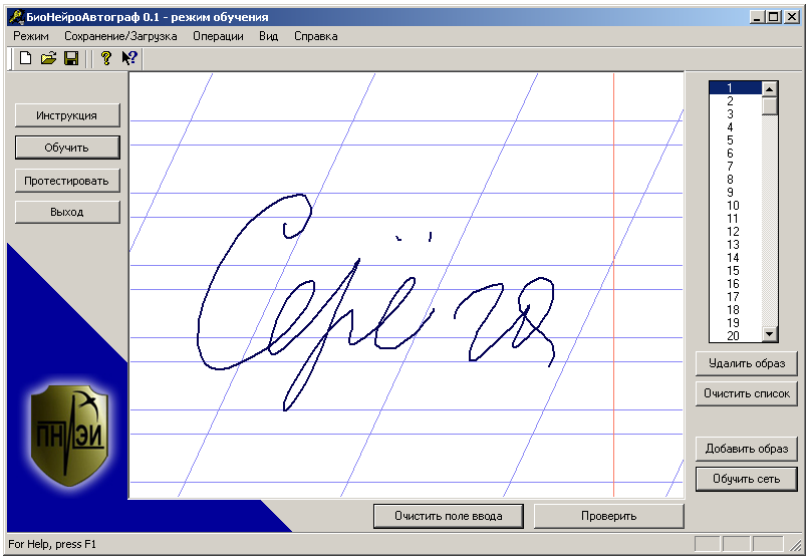
16. Удалите все обучающие примеры из списка, нажав кнопку "Очистить список".

17. Создайте тестовую базу образов "Чужие". Тестовая база должна содержать от 128 до 256 различных рукописных слов. Ввод и добавление тестовых образов в список осуществляется аналогично вводу обучающих образов.

18. Сохраните сформированную тестовую базу (пункт меню "Сохранение/Загрузка" подпункт "Сохранить образы на диск").

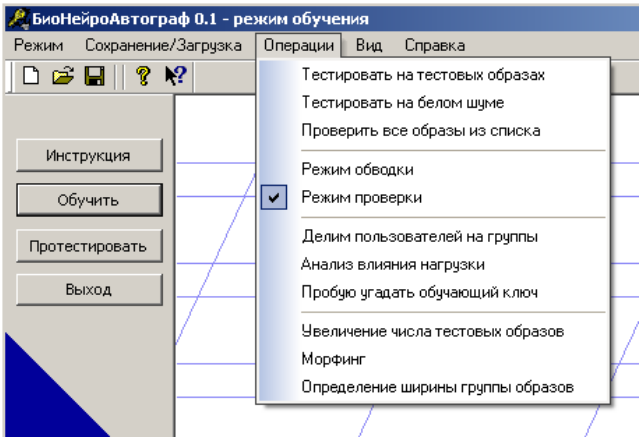


Допускается использование готовой базы образов "все Чужие", созданной совместными усилиями нескольких человек. Загрузка сформированной ранее базы образов осуществляется с помощью подпункта "Загрузить образы с диска" меню "Сохранение/Загрузка". В появившемся диалоговом окне необходимо выбрать файл с образами, например, "256_ВСЕ_ЧУЖИЕ.dat". Загруженные примеры рукописных слов отображаются в списке.

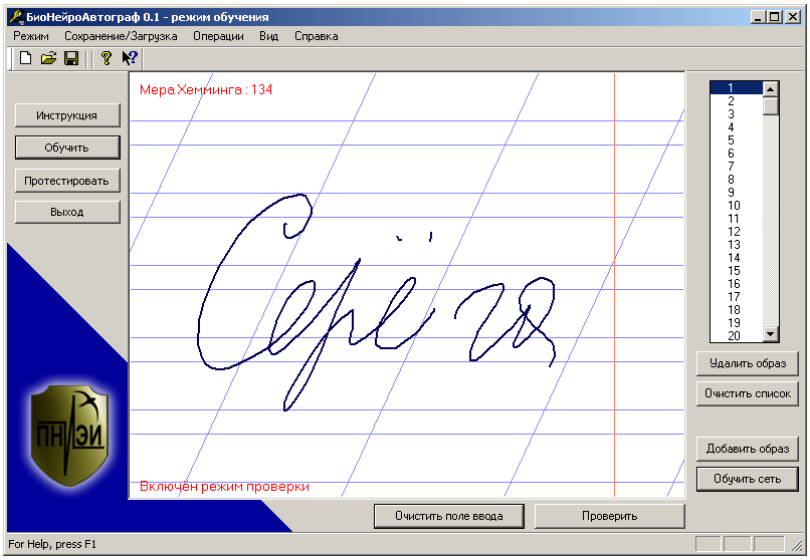


19. Выберите один из загруженных примеров, щёлкнув по его номеру "мышкой" в списке образов. Проверьте насколько выбранный образ отличается от обучающих, нажав на кнопку "Проверить".

20. Для ускоренного тестирования обученной нейронной сети на загруженных образах выберите режим проверки, установив галочку возле подпункта "Режим проверки" пункта меню "Операции".



В этом режиме вы видите расстояние/меру Хемминга между обучающим ключом и ключом, полученном на тестовом образе. Чем сильнее тестовый образ отличается от обучающего, тем больше мера Хемминга. На обучающих образах мера равна нулю.

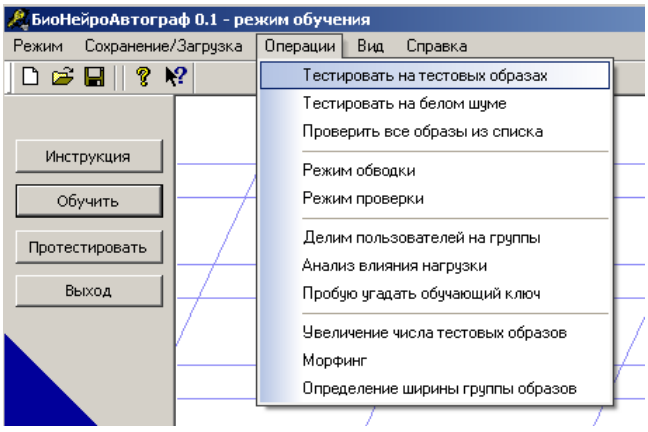


21. Протестируйте обученную нейронную сеть на всех тестовых образах и сформируйте таблицу попадания расстояний Хемминга в разные интервалы, факт попадания меры Хемминга в определённый интервал отмечайте символом '*'. Например, на рисунке выше мера Хемминга на образе "Серёга" равна 134, следовательно '*' ставится в интервал 130 – 140 бит. После окончания тестирования подсчитывается количество '*' в разных интервалах и заполняется столбец "Число попаданий" (таблица 1).

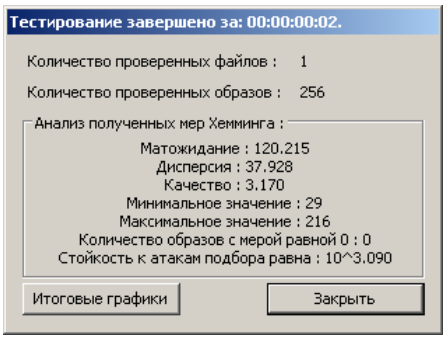
Таблица 1

Интервал расстояний	Отметки факта попадания в интервал	Число попаданий
20 – 30 бит	*	n_1=1
30 – 40 бит	*******	n_2=7
40 – 50 бит	*	n_3=1
50 – 60 бит	**********	n_4=11
60 – 70 бит	*****	n_5=5
70 – 80 бит	**********	n_6=16
80 – 90 бит	**********	n_7=15
90 – 100 бит	**********	n_8=19
100 – 110 бит	**********	n_9=30
110 – 120 бит	**********	n_{10}=22
120 – 130 бит	**********	n_{11}=32
130 – 140 бит	**********	n_{12}=23
140 – 150 бит	**********	n_{13}=16
150 – 160 бит	**********	n_{14}=11
160 – 170 бит	**********	n_{15}=14
170 – 180 бит	**********	n_{16}=9
180 – 190 бит	*******	n_{17}=7
190 – 200 бит	*****	n_{18}=5
200 – 210 бит	**	n_{19}=2
210 – 220 бит	***	n_{20}=3
		$\sum n_i = N = 256$

22. Вычислите математическое ожидание и среднеквадратическое отклонение расстояний Хэмминга. Для этого необходимо запустить тестирование на тестовых образах (пункт меню "Операции" подпункт "Тестировать на тестовых образах") и в появившемся окне указать путь к файлу с тестовыми образами.



После завершения тестирования выводится диалоговое окно с результатами.



Из экранной формы видно, что значение математического ожидания меры Хемминга E(h) на тестовых образах равно 120.215, а среднеквадратическое отклонение σ(h) равно 37.928.

23. Вычислите среднюю вероятность попадания в выделенные ранее интервалы для нормального закона распределения значений со значениями первых статистических моментов E(h) = 120.215, σ(h) = 37.928 по следующей формуле:

$$p = \frac{1}{\sigma(h)\sqrt{2\pi}} \cdot \int_{left}^{right} e^{-\frac{(x-E(h))^2}{2\sigma(h)^2}} \cdot dx, \qquad (1)$$

где *left* – левая граница рассматриваемого интервала;

right – правая граница рассматриваемого интервала.

Полученные данные сведите в таблицу 2.

Таблица 2.

Интервал расстояний	Теоретическая вероятность попадания в интервал
20 – 30 бит	p_1= 0.0045585
30 – 40 бит	p_2= 0.00851081
40 – 50 бит	p_3= 0.01482712
50 – 60 бит	p_4= 0.02410355
60 – 70 бит	p_5= 0.03656314
70 – 80 бит	p_6= 0.05175414
80 – 90 бит	p_7= 0.06835748
90 – 100 бит	p_8= 0.08424939
100 – 110 бит	p_9= 0.09689191
110 – 120 бит	p_{10}= 0.10397968
120 – 130 бит	p_{11}= 0.10412372
130 – 140 бит	p_{12}= 0.09729513
140 – 150 бит	p_{13}= 0.08483455
150 – 160 бит	p_{14}= 0.0690231
160 – 170 бит	p_{15}= 0.05240297
170 – 180 бит	p_{16}= 0.03712417
180 – 190 бит	p_{17}= 0.02454125
190 – 200 бит	p_{18}= 0.01513823
200 – 210 бит	p_{19}= 0.00871347
210 – 220 бит	p_{20}= 0.00467999

24. Вычислите значение хи-квадрат с 20 степеням свободы:

$$\chi^2 = \sum_{i=1}^{20} \frac{(n_i - N \cdot p_i)^2}{N \cdot p_i} = 2.91,$$

где N – число выполненных опытов (N=256);

n_i – число попаданий в i-й интервал;

p_i – теоретическая вероятность попадания в интервал, если гипотеза нормальности распределения верна.

25. Вычислите вероятность того, что гипотеза нормальности для полученных данных выполняется:

$$P = 1 - \alpha(17, 2.91) = 1 - \int_0^{2.91} p\left(\chi^2(17, u)\right) du = 0.99986,$$

где α – кантиль хи-квадрат распределения;

$p(\chi^2(17, u))$ – плотность хи-квадрат распределения значений с 17 степенями свободы.

Выбрано число степеней свободы (17 = 20-1-2), уменьшенное в сравнении с обычной оценочной величиной (20-1) из-за того, что 2 момента (математическое ожидание и среднеквадратическое отклонение в формуле (1)) вычислялись на той же самой выборке исходных данных.

ВЫВОД: гипотеза нормальности полученного распределения значений расстояний Хемминга верна с вероятностью 0.99986 по критерию хи-квадрат с 17 степенями свободы. Стойкость к атакам подбора тестируемого нейросетевого преобразователя биометрия-код (10^3) сопоставима с обратной величиной квантиля хи-квадрат распределения, используемого при проверке гипотезы.