

Supplementary Materials:

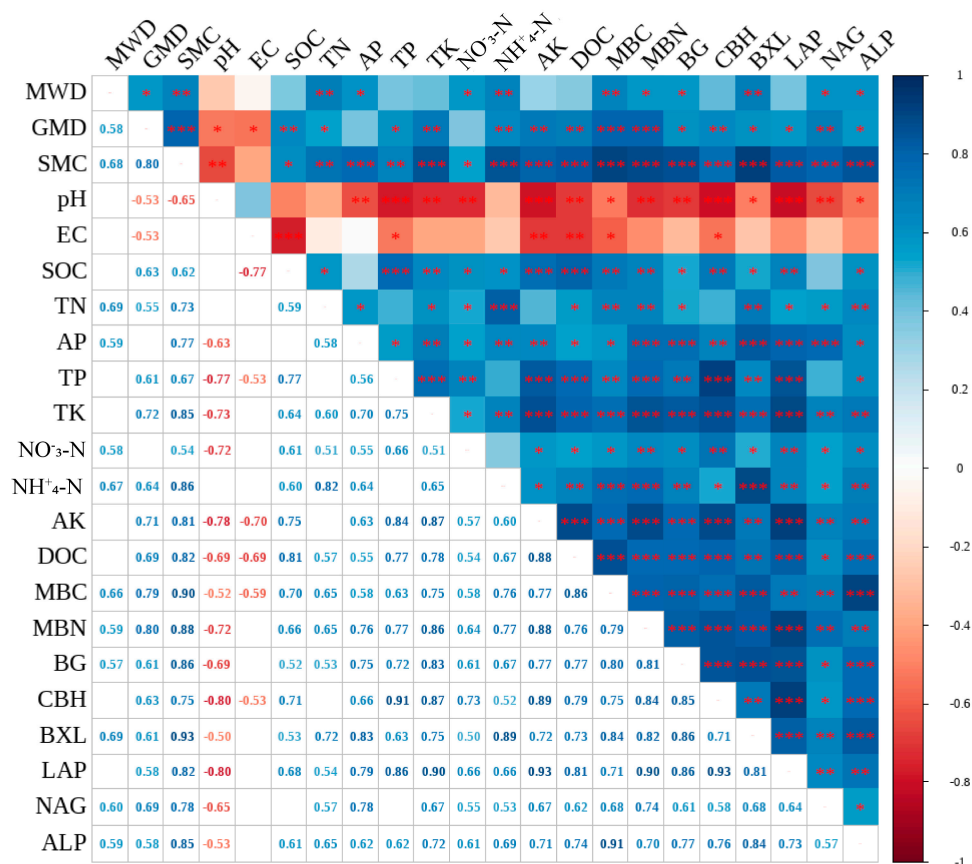


Figure S1. The correlation heatmap of soil indicators from total data set. MWD mean weight diameter, GMD geometric mean diameter, SMC soil moisture content (%), EC electrical conductivity (us cm⁻¹), SOC soil organic carbon (g kg⁻¹), TN total nitrogen (g kg⁻¹), AP available phosphorus (mg kg⁻¹), TP total phosphorus (g kg⁻¹), TK total potassium (mg kg⁻¹), NO₃-N nitrate nitrogen (mg kg⁻¹), NH₄⁺-N ammonium nitrogen (mg kg⁻¹), AK available potassium (mg kg⁻¹), DOC dissolved organic carbon (mg kg⁻¹), MBC microbial biomass carbon (mg kg⁻¹), MBN microbial biomass nitrogen (mg kg⁻¹), MBN ratio of microbial biomass nitrogen to total nitrogen, BG β-glucosidase (IU g⁻¹), CBH β-cellobiosidase (IU g⁻¹), BXL β-Xylosidase (IU g⁻¹), NAG β-1,4-N-acetylglucosaminidase (U g⁻¹), LAP L-leucine aminopeptidase (U g⁻¹), ALP Alkaline Phosphatase (IU g⁻¹).

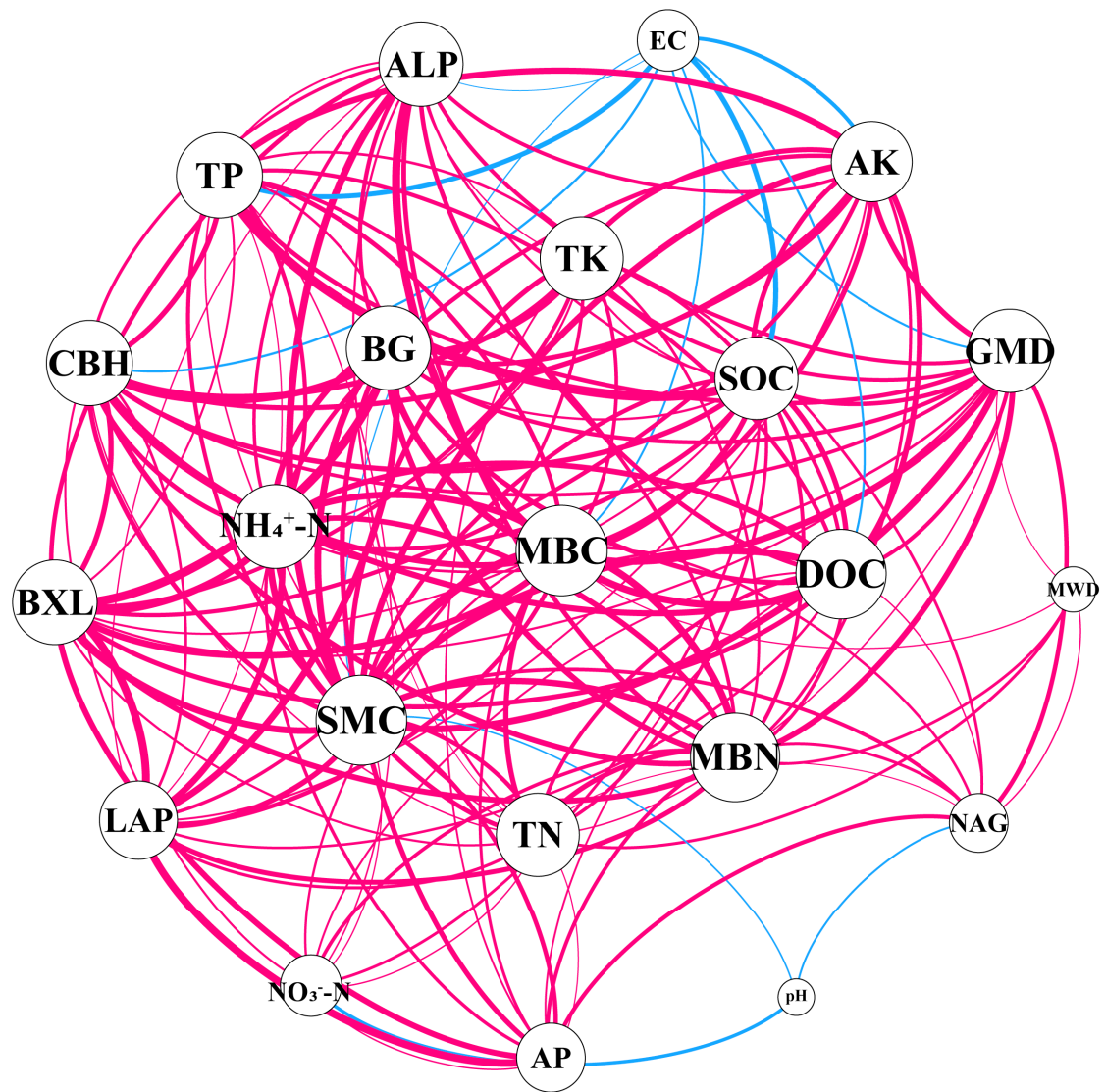


Figure S2. The network model generated using 22 soil parameters. Each node (circle) represents a soil parameters and edge (line) shows the significant correlations between soil parameters across different treatments. The size of node reflects the proportional to the eigenvector centrality. The thickness of each line among two nodes is proportional to the value of Spearman's coefficient $>|\pm 0.60|$ and $P < 0.05$ to reach the bivariate correlations. The size and color density of the lines change to show that the varying strength of relationship among soil parameters. In addition, red and blue line reflect positivity correlation and negative between two nodes, respectively. See Table 1 for abbreviation meanings.

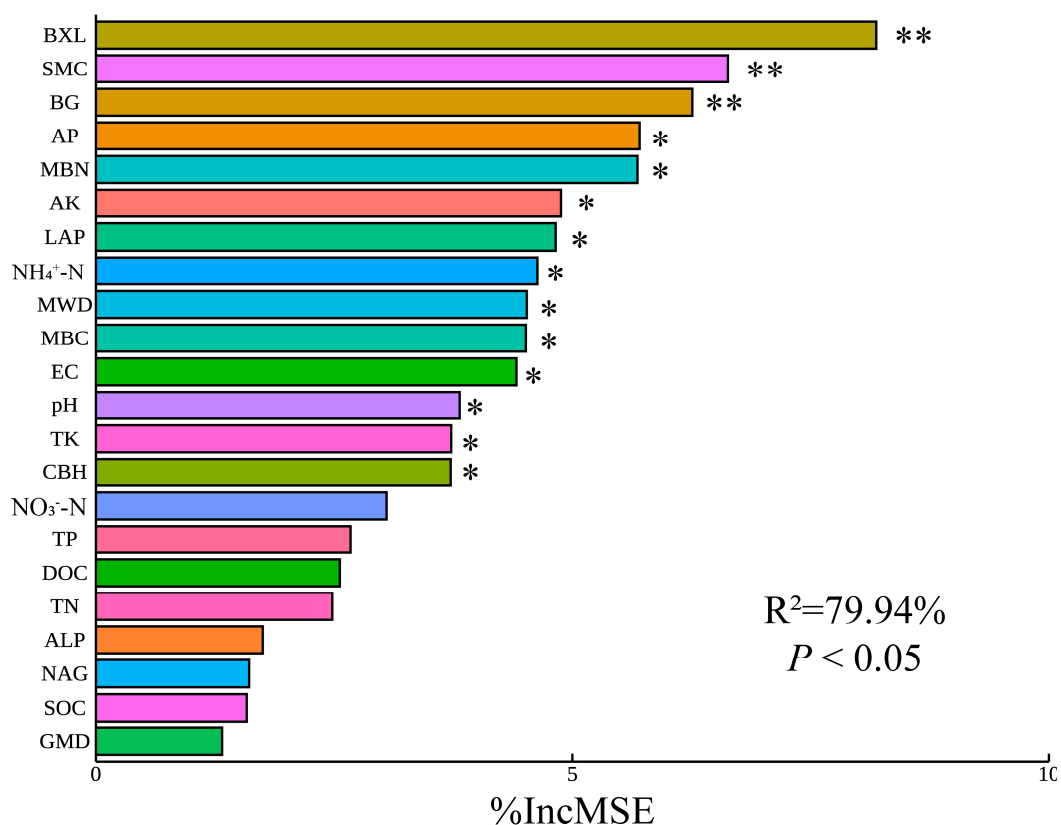


Figure S3. Random forest analysis means predictor importance (%IncMSE: % of increase of MSE) of soil parameters on millet yield. significance level of each soil parameters is as follow: * $P < 0.05$, ** $P < 0.01$. See Table 1 for abbreviation meanings.

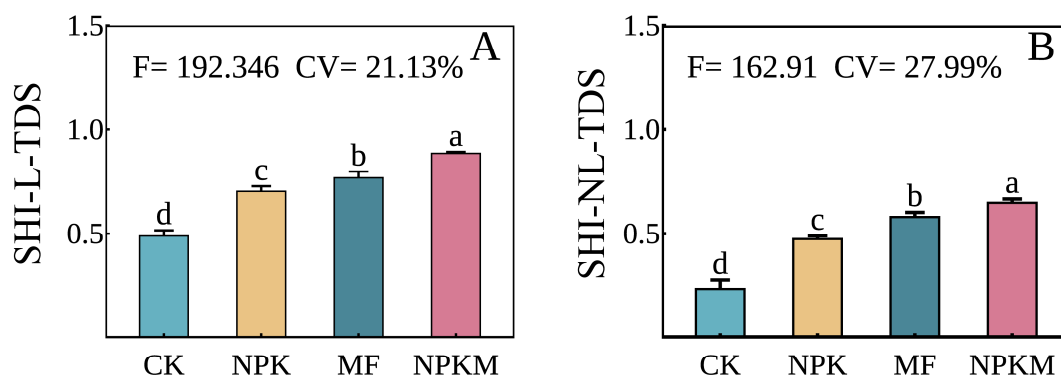


Figure S4. Soil health indexes (SHIs) under four treatments. Values are means \pm standard error ($n = 4$). Different letters indicate significant difference among treatments ($P < 0.05$). SHI-L-TDS, soil health index calculated by linear scoring curve and total data set; SHI-NL-TDS, soil health index calculated by nonlinear scoring curve and total data set.

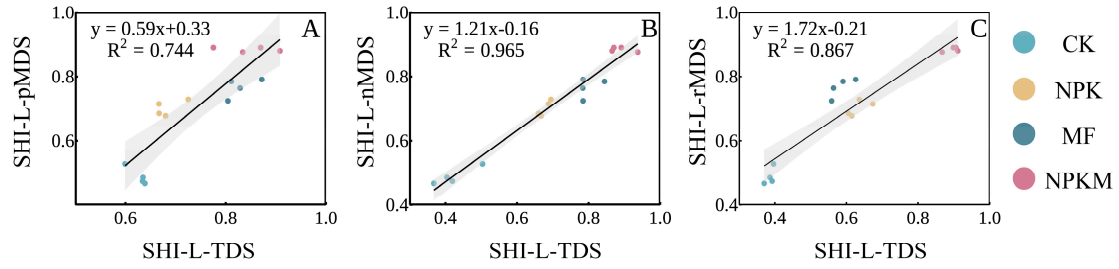


Figure S5. Relationship between of soil health index based on three minimum data sets between total data set methods under linear scoring curve. SHI-L-pMDS, soil health index calculated by linear scoring curve and minimum data set established with principal component analysis; SHI-L-nMDS, soil health index calculated by linear scoring curve and minimum data set established with network analysis; SHI-L-rMDS, soil health index calculated by linear scoring curve and minimum data set established with random forest analysis; SHI-L-TDS, soil health index calculated by linear scoring curve and total data set.

Table S1 Field fertilization management description of the different treatment from this study in the agro-ecotone of Northern China.

Treatments	Specific operation	Millet phase			
		Basal fertilization(kg·ha ⁻¹)			
		N	P ₂ O ₅	K ₂ O	Cow manure
CK	No fertilization	0	0	0	0
NPK	Only chemical fertilizer is applied	240	96	153.6	0
MF	Only cow manure is applied	0	96	153.6	1600
NPKM	50% chemical fertilizer plus 50% cow manure is applied	120	96	153.6	800

Table S2. Descriptive statistics and measurement methods of soil physical, chemical, and biological properties.

Soil property	Abbreviation	Unit	Method description
Mean weight diameter	MWD	mm	Wet sieving method (He et al., 2020)
Geometric mean diameter	GMD	mm	Wet sieving method (He et al., 2020)
Soil moisture content	SMC	%	Weighted and dried at 105 °C for 24 h
pH	pH	Unitless	The ratio of soil / water with 1:2.5 using a pH meter (FE28, Mettler Toledo, Switzerland) (Lu 1999)
Electrical conductivity	EC	Us cm ⁻¹	The ratio of soil / water with 1:5 using a EC meter (FE38, Mettler Toledo, Switzerland) (Lu 1999)
Soil organic carbon	SOC	g kg ⁻¹	K ₂ Cr ₂ O ₇ colorimetric oxidization method (Lu 1999)
Total nitrogen	TN	g kg ⁻¹	Kjeldahl method (Lu 1999)
Available phosphorus	AP	mg kg ⁻¹	NaHCO ₃ extraction (0.5 mol·L ⁻¹) and spectrophotometrically determined at 880 nm using the phosphomolybdate method (Olsen et al., 1954)
Total phosphorus	TP	g kg ⁻¹	HF-HClO ₄ extraction and molybdenum blue colorimetric method (Lu 1999)
Total potassium	TK	mg kg ⁻¹	HNO ₃ -HClO ₄ extraction and measured by atomic spectrophotometry (Lu 1999)
Available potassium	AK	mg kg ⁻¹	Ammonium (1 mol·L ⁻¹) acetate and determined by flame emission spectrophotometry (Olsen et al., 1954)
Dissolved organic carbon	DOC	mg kg ⁻¹	Deionized water extraction (soil / water 1:5) and measured with Elementar Vario EL III TOC analyzer (Lu 1999)
Nitrate nitrogen	NO ₃ -N	mg kg ⁻¹	KCl extraction (2 mol·L ⁻¹) detected on a continuous flow analytical system (Skalar Analytical, Breda, the Netherlands) (Lu 1999)
Ammonium nitrogen	NH ₄ ⁺ -N	mg kg ⁻¹	
Microbial biomass carbon	MBC	mg kg ⁻¹	Chloroform fumigation - incubation method (Wu et al., 1990)
Microbial biomass nitrogen	MBN	mg kg ⁻¹	
β-1,4-Glucosidase	BG	IU g ⁻¹	Double antibody sandwich method (Marx et al., 2001)
β-Xylosidase	BXL	IU g ⁻¹	

β - Cellobiohydrolase	CBH	IU g ⁻¹
L-leucine aminopeptidase	LAP	U g ⁻¹
β -1,4-N-Acetyl-glucosaminidases	NAG	U g ⁻¹
Alkaline phosphatase	ALP	IU g ⁻¹

Table S3. Results of principal component analysis of total data set (eigenvalues, eigenvectors and the percentage of total variance explained by each extracted factor). .

PCs	PC1	PC2	PC3	COM
MWD	0.169	0.85	0.002	0.751
GWD	0.294	0.562	0.548	0.703
SMC	0.519	0.755	0.33	0.949
pH	<u>-0.893</u>	-0.104	-0.243	0.867
EC	-0.192	0.028	<u>-0.933</u>	0.907
SOC	0.302	0.32	0.813	0.854
TN	0.215	0.814	0.165	0.736
AP	0.735	0.595	-0.177	0.926
TP	0.719	0.227	0.499	0.818
TK	0.7	0.458	0.354	0.825
NO ₃ -N	0.641	0.288	0.243	0.553
NH ₄ ⁺ -N	0.201	<u>0.865</u>	0.282	0.868
AK	0.714	0.293	0.586	0.938
DOC	0.528	0.41	0.654	0.876
MBC	0.361	0.693	0.547	0.909
MBN	0.653	0.571	0.361	0.884
BG	0.693	0.54	0.229	0.824
CBH	0.796	0.297	0.445	0.921
BXL	0.484	0.793	0.195	0.901
LAP	0.82	0.376	0.352	0.938
NAG	0.568	0.568	0.101	0.655
ALP	0.431	0.635	0.405	0.753
Eigenvalue	14.87	2.09	1.40	
Variance explained (%)	67.59	9.50	6.35	
Cumulative variance (%)	67.59	77.10	83.44	

Note: PC principal component, COM communalities. Bold and underlined loading values correspond to the soil properties included in the PCA-MDS (pMDS). See Table 1 for abbreviation definitions.

Table S4. The linear scoring methods (L) soil health index (SHI) models based on the principal component analysis (PCA), random forest analysis (RF) and network analysis (NA) methods.

SHI model		Mean	Max.	Min.
SHI-L-pMDS	pH×0.328+ EC×0.343+	0.75	0.94	0.37
SHI-NL-pMDS	NH ₄ ⁺ -N×0.329	0.49	0.62	0.34
SHI-L-nMDS	GMD×0.133+ SMC×0.149+ TP×0.138+ DOC×+0.148	0.70	0.94	0.37
SHI-NL-nMDS	MBC×0.150+ MBN×0.145+ CBH×0.138	0.48	0.67	0.19
SHI-L-rMDS	MWD×0.266+ SMC×0.216+ AP×0.186+	0.62	0.91	0.37
SHI-NL-rMDS	MBN×0.185+ BXL×0.266	0.47	0.72	0.21

Note: SHI-L-pMDS, linear scoring curve and based on principal component analysis to establish minimum data set; SHI-NL-pMDS, nonlinear scoring curve and based on principal component analysis to establish minimum data set; SHI-L-nMDS, linear scoring curve and based on network analysis to establish minimum data set; SHI-NL-nMDS, nonlinear scoring curve and based on network analysis to establish minimum data set; SHI-L-rMDS, linear scoring curve and based on random forest analysis to establish minimum data set; SHI-NL-rMDS, nonlinear scoring curve and based on random forest analysis to establish minimum data set; See Table 1 for abbreviation definitions.

Table S5. Correlation matrix for the eight soil health indexes and millet yield.

	SHI-L-TDS	SHI-L-pMDS	SHI-L-nMDS	SHI-L-rMDS	SHI-NL-TDS	SHI-NL-pMDS	SHI-NL-nMDS	SHI-NL-rMDS	Yield
SHI-L-TDS	1								
SHI-L-pMDS	0.863**	1							
SHI-L-nMDS	0.983**	0.857**	1						
SHI-L-rMDS	0.931**	0.731**	0.872**	1					
SHI-NL-TDS	0.999**	0.858**	0.982**	0.924**	1				
SHI-NL-pMDS	0.882**	0.994**	0.873**	0.751**	0.878**	1			
SHI-NL-nMDS	0.982**	0.859**	0.998**	0.865**	0.984**	0.874**	1		
SHI-NL-rMDS	0.946**	0.738**	0.893**	0.995**	0.943**	0.758**	0.890**	1	
Yield	0.930**	0.712**	0.892**	0.954**	0.926**	0.731**	0.889**	0.956**	1

Note: SHI-L-TDS, linear scoring curve and total data set; SHI-NL-TDS, nonlinear scoring curve and total data set; SHI-L-pMDS, linear scoring curve and based on principal component analysis to establish minimum data set; SHI-NL-pMDS, nonlinear scoring curve and based on principal component analysis to establish minimum data set; SHI-L-nMDS, linear scoring curve and based on network analysis to establish minimum data set; SHI-NL-nMDS, nonlinear scoring curve and based on network analysis to establish minimum data set; SHI-L-rMDS, linear scoring curve and based on random forest analysis to establish minimum data set; SHI-NL-rMDS, nonlinear scoring curve and based on random forest analysis to establish minimum data set. ** Significant at the 0.01 level. *Significant at the 0.05 level.